Weakly Supervised Regression with Interval Targets

Xin Cheng [1]   **Yuzhou Cao** [2]   **Ximing Li** [3]   **Bo An** [2]   **Lei Feng** [2]

Abstract

This paper investigates an interesting weakly supervised regression setting called *regression with interval targets* (RIT). Although some of the previous methods on relevant regression settings can be adapted to RIT, they are not *statistically consistent*, and thus their empirical performance is not guaranteed. In this paper, we provide a thorough study on RIT. First, we proposed a novel statistical model to describe the data generation process for RIT and demonstrate its validity. Second, we analyze a simple selection method for RIT, which selects a particular value in the interval as the target value to train the model. Third, we propose a statistically consistent limiting method for RIT to train the model by limiting the predictions to the interval. We further derive an estimation error bound for our limiting method. Finally, extensive experiments on various datasets demonstrate the effectiveness of our proposed method.

1. Introduction

Regression is a significantly important task in machine learning and statistics (Stulp & Sigaud, 2015; Uysal & Güvenir, 1999). The goal of the regression task is to learn a predictive model from a given set of training examples, where each training example consists of an instance (or feature vector) and a *real-valued target*. Conventional supervised regression normally requires a vast amount of labeled data to learn an effective regression model with excellent performance. However, it could be difficult to obtain *fully supervised* training examples due to the high cost of data labeling in real-world applications. To alleviate this problem, many weakly supervised regression settings have been investigated, such as semi-supervised regression (Li et al., 2017;

Fully supervised regression   Regression with interval targets

The true age: 58.   The age interval: [40, 70].

Figure 1. An example of facial age estimation. Both photos were taken when Pele was 58 years old.

Wasserman & Lafferty, 2007; Kostopoulos et al., 2018), multiple-instance regression (Amar et al., 2001; Wang et al., 2011; Park et al., 2020), uncoupled regression (Carpentier & Schlüter, 2016; Xu et al., 2019), and regression with noisy targets (Ristovski et al., 2010; Hu et al., 2020).

This paper investigates another interesting weakly supervised regression setting called *regression with interval targets* (RIT). For RIT, we aim to learn a regression model from weakly supervised training examples, each annotated with *only an interval that contains the true target value*. The learned regression model in this setting is expected to predict the target value of any test instance as accurately as possible. In many real-world scenarios, it is difficult to collect the exact true target value, while it could be easy to provide an interval in which the true target value is contained. A typical example is facial age estimation (Geng et al., 2013). In Figure 1, there are two photos of Ballon d'Or King Pele at the age of 58. The fully supervised regression task requires the exact age of Pele, which is quite difficult to provide because it is common for a person to look the same over a long period of time. However, we can easily get an age interval that contains the true age of Pele. Based on the facial wrinkles, we can determine that the true age is at least 40 but not more than 70. In reality, many regression tasks face this challenge (especially in size/length/age estimation), where it is costly or impossible to obtain a true target value.

Our studied RIT is highly related to *interval-valued data prediction* (IVDP) (Ishibuchi & Tanaka, 1991; Neto & De Carvalho, 2008; 2010; Fagundes et al., 2014). IVDP allows

---

[1]College of Computer Science, Chongqing University, China [2]School of Computer Science and Engineering, Nanyang Technological University, Singapore [3]College of Computer Science and Technology, Jilin University, China. Correspondence to: Lei Feng <lfengqaq@gmail.com>.

each training example to be annotated with an interval and learns a regression model to *predict the interval* containing the true target value of a test instance. There are even some studies that allow features to be intervals as well (Manski & Tamer, 2002; Yang & Liu, 2018; Sadeghi et al., 2019). It is worth noting that IVDP aims to predict the interval that contains the target value, while our studied RIT aims to predict the true target value. Although some of the above methods could be adapted to our studied RIT settings, they are not *statistically consistent* (i.e., the learned model is infinite-sample consistent to the optimal model), and thus the empirical performance is not guaranteed.

In this paper, we provide a thorough study on RIT, and the main contributions can be summarized as follows:

- We propose a novel statistical model to describe the data generation process for RIT and demonstrate its validity. Having an explicit data distribution helps us understand how data with interval targets are generated.

- We analyze a simple selection method for RIT, which selects a particular value in the interval as the target value to train the model. We show that this intuitive method could work well if the middlemost value in the interval is taken as the target value.

- We propose a *statistically consistent* limiting method for RIT to train the model by limiting the predictions of the model to the interval. We further derive an *estimation error bound* for our limiting method.

Extensive experiments on various datasets demonstrate the effectiveness of our proposed method.

2. Related Work

Regression. For the ordinary regression problem, let the feature space be $\mathcal{X} \in \mathbb{R}^d$ and the label space be $\mathcal{Y} \in \mathbb{R}$. Let us denote by $(\boldsymbol{x}, y)$ an example including an instance x and a real-valued true label y. Each example $(\boldsymbol{x}, y) \in \mathcal{X} \times \mathcal{Y}$ is assumed to be independently sampled from an unknown data distribution with probability density $p(\boldsymbol{x}, y)$. For the regression task, we aim to learn a model $f : \mathcal{X} \mapsto \mathbb{R}$ that tries to minimize the following expected risk:

$$R(f) = \mathbb{E}_{p(\boldsymbol{x}, y)}[\ell(f(\boldsymbol{x}), y)], \qquad (1)$$

where $\mathbb{E}_{p(\boldsymbol{x}, y)}$ denotes the expectation over the distribution $p(\boldsymbol{x}, y)$ and $\ell : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_+$ is a conventional loss function (such as mean squared error and mean absolute error) for regression, which measures how well a model estimates a given real-valued label.

Interval-valued data prediction. In order to consider multiple types of data, such as intervals, weights, and characters, symbolic data analysis (Bock & Diday, 1999; Billard, 2006) has been extensively investigated. As a specific task of

symbolic data analysis, the purpose of *interval-valued data prediction* (IVDP) is to learn an interval predictor from training data annotated with intervals. The challenge of IVDP is mainly to construct a model that outputs intervals (to ensure that the interval holds, e.g., $[0, 10]$ instead of $[10, 0]$). In statistics, Billard & Diday (2000) introduced a central tendency for interval data. (Lauro & Palumbo, 2000) introduced principal component analysis methods for interval data. In addition, by designing proper loss functions, network structures, or output constraints, neural networks can also be trained to output intervals for IVDP (Neto & De Carvalho, 2008; 2010; Giordani, 2015; Yang et al., 2019; Sadeghi et al., 2019).

Regression with interval-censored data. Another related setting is *regression with interval-censored data* (RICD) (Rabinowitz et al., 1995; Lindsey & Ryan, 1998; Lesaffre et al., 2005; Sun, 2006), which aims to learn a survival function from interval-censored data. In a sequence of time points, a specific event (e.g., machine breakdown, disease attack, death) occurs between two time points, and the interval formed by these two time points is called interval censoring. RICD was widely used in survival analysis (Machin et al., 2006; Kleinbaum et al., 2012; Wang et al., 2019b). In contrast to our studied RIT setting, RICD aims to obtain a survival function to estimate the occurring probability of an event, instead of learning a predictive model.

3. Regression with Interval Targets

Notations. Suppose the given training set is denoted by $\{(\boldsymbol{x}_i, S_i)\}_{i=1}^n$ where S_i represents the interval $[\underline{y}_i, \overline{y}_i] \in \mathcal{Y} \times \mathcal{Y}$ assigned to the instance $\boldsymbol{x} \in \mathcal{X}$, and $|S_i| = \overline{y}_i - \underline{y}_i$ represents the size of the interval S_i. Each training example $(\boldsymbol{x}_i, S_i)$ is assumed to be sampled from an unknown joint distribution with probability density $\tilde{p}(\boldsymbol{x}, S)$. In this setting, the true label $y_i \in \mathcal{Y}$ of the instance $\boldsymbol{x}_i$ is guaranteed to be contained in the interval S_i. The goal of interval regression is to induce a regression model $f : \mathcal{X} \mapsto \mathbb{R}$ that can accurately predict the target value of a test instance. Interestingly, this setting can be considered as a generalized setting of ordinary regression, because we can easily convert the ordinary regression example $(\boldsymbol{x}_i, y_i)$ to an interval regression example by rewriting y_i as the interval $[\underline{y}_i, \overline{y}_i]$ where $\underline{y}_i = \overline{y}_i = y_i$. In this paper, we denote by $p(\cdot)$ the probability density and $\Pr[\cdot]$ the occurring probability.

Small ambiguity degree. For ensuring that RIT is learnable (i.e., the true target value concealed in the interval is distinguishable), we assume that RIT should satisfy the small ambiguity degree condition (Cour et al., 2011), where the ambiguity degree in our setting is defined as

$$\lambda = \sup_{(\boldsymbol{x}, y) \sim p(\boldsymbol{x}, y), (\boldsymbol{x}, S) \sim \tilde{p}(\boldsymbol{x}, S), y' \in \mathcal{Y}, y' \neq y} \Pr[y' \in S].$$

The ambiguity degree λ is the maximum probability of a specific incorrect target y' co-occurring with the true target y in the same interval S. We can observe that when $\lambda = 1$, the incorrect target y' always appears with the true target y together, and thus we can no longer distinguish which one is the true target. Therefore, the RIT setting requires to assume that the small ambiguity degree condition is satisfied (i.e., $\lambda < 1$), in order to ensure that this setting is learnable.

3.1. Data Generation Process

To avoid the sampled intervals being unreasonable (i.e., the size is unexpectedly large), we use q to denote the maximum allowed interval size $|S|$. Then, we assume that each example $(\boldsymbol{x}, S)$ with $S = [\underline{y}, \overline{y}]$ is independently sampled from a probability distribution with the following density:

$$\tilde{p}(x, S) = \int_{\underline{y}}^{\overline{y}} p(S|y)p(\boldsymbol{x}, y)\mathrm{d}y, \tag{2}$$

where

$$p(S|y) = \begin{cases} \dfrac{2}{q^2}, & y \in S \text{ and } |S| \leq q, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

In Eq. (2), we assume $p(S|y, \boldsymbol{x}) = p(S|y)$, which means that given the true label y, the interval S is independent of the instance $\boldsymbol{x}$. Such a class-dependent and instance-independent assumption was widely adopted by many previous studies in the weakly supervised learning field (Patrini et al., 2017; Ishida et al., 2019; Feng et al., 2020). In Eq. (3), we assume that given a specific label y, all possible intervals $\{S | y \in S, |S| \leq q\}$ are uniformly sampled.

We show that our presented joint distribution $\tilde{p}(\boldsymbol{x}, S)$ is a valid probability distribution by the following theorem.

Theorem 3.1. *The following equality holds:*

$$\int_S \int_{\mathcal{X}} \tilde{p}(\boldsymbol{x}, S)\mathrm{d}\boldsymbol{x}\mathrm{d}S = 1. \tag{4}$$

In addition to proving that $\tilde{p}(\boldsymbol{x}, S)$ is a valid probability distribution, we also need to verify that $\tilde{p}(\boldsymbol{x}, S)$ meets the key requirement of RIT, i.e., the true target y is guaranteed to be contained in the interval S for every example $(\boldsymbol{x}, S)$ sampled from $\tilde{p}(\boldsymbol{x}, S)$. The following theorem provides an affirmative answer to this question.

Theorem 3.2. *For any interval example $(\boldsymbol{x}, S)$ independently sampled from the assumed data distribution $\tilde{p}(\boldsymbol{x}, S)$ defined in Eq. (2), the true target y is always in the interval S, i.e., $\Pr[y \in S | \boldsymbol{x}, S] = 1, \forall(\boldsymbol{x}, S) \backsim \tilde{p}(\boldsymbol{x}, S)$.*

3.2. Real-World Motivation

Here, we give a real-world motivation for the assumed data distribution $\tilde{p}(\boldsymbol{x}, S)$. For the data annotations in the regression task, it could be difficult to directly provide the exact

true target value for each instance. Fortunately, it would be easier if the annotation system can randomly generate an interval and ask annotators whether the true target value is contained in the generated interval or not. Given an instance $\boldsymbol{x}$, the maximum size q of the interval, and the maximum and minimum values $y_{\max}, y_{\min}$ of the label space $\mathcal{Y}$, suppose the annotation system randomly and uniformly samples $\underline{y}$ and $\overline{y}$ from the interval $[y_{\min} - q, y_{\max} + q]$ ($\pm q$ is to ensure that all possible y have the same possible number of intervals) to generate an interval $S = [\underline{y}, \overline{y}]$. If the sampled $\underline{y}$ and $\overline{y}$ satisfy the two conditions: $\overline{y} - \underline{y} \leq q$ and the true target value y of $\boldsymbol{x}$ belongs to $[\underline{y}, \overline{y}]$ i.e., $y \in [\underline{y}, \overline{y}]$, then we collect an interval regression example $(\boldsymbol{x}, S)$ where $S = [\underline{y}, \overline{y}]$, otherwise we discard the interval S for the instance $\boldsymbol{x}$. In this way, each collected interval regression example $(\boldsymbol{x}, S)$ exactly follows the data distribution defined in Eq. (2). We will demonstrate this argument below.

We start by considering the case where the annotation system has discarded all intervals larger than a given maximum interval value q. Then we have the following lemma.

Lemma 3.3. *Given the maximum value q allowed for the interval size, and the maximum value $y_{\max}$ and the minimum value $y_{\min}$ of the label space $\mathcal{Y}$, for any instance $\boldsymbol{x}$ with its true target y and any interval S with size no greater than q (i.e., $|S| \leq q$), the following equality holds:*

$$\Pr[y \in S | \boldsymbol{x}] = \frac{q}{2(y_{\max} - y_{\min}) + q}. \tag{5}$$

In the case of no additional information, we can only choose the interval randomly, so the above probability is uniform. When the maximum value q allowed in the interval increases, the probability in Eq. (5) will increase, which is in line with our knowledge because a larger interval is more likely to contain the true value y. Similarly, the larger the space $(y_{\max} - y_{\min})$ allowed for sampling, the more difficult it is to obtain an interval containing the true label y. Based on lemma 3.3, we have the following theorem.

Theorem 3.4. *Under the same setting of Lemma 3.3, the distribution of collected data where the true label $y \in \mathcal{Y}$ of an example $\boldsymbol{x}$ belongs to the interval S is the same as Eq. (2), i.e.,*

$$p(x, S | y \in S) = \tilde{p}(x, S).$$

Theorem 3.4 clearly demonstrates that our assumed data distribution $\tilde{p}(\boldsymbol{x}, y)$ exactly accords with the real-world motivation introduced above.

4. The Proposed Methods

In this section, we introduce a simple method that selects a *particular value* in the interval as the target value to train a

regression model. This method is simple and intuitive, but it only considers a single value in the interval and ignores the overall interval information. To overcome this drawback, we propose a limiting method that limits the predicted value of the model to be in the interval, which is *statistically consistent* under a very mild condition.

4.1. The Simple Selection Method

Given an interval, an intuitive solution is to select a particular value in the interval as the target value:

$$\ell_{\text{sel}}(f(\boldsymbol{x}), S) = \ell(f(\boldsymbol{x}), y'), \text{where } y' \in S. \quad (6)$$

As shown in Eq. (6), this method aims to select one value in the interval as the target value and regard the loss of this value as the predictive loss for the interval regression example $(\boldsymbol{x}, S)$. This simple method has an obvious drawback, i.e., selecting only one value in the interval ignores the influence of other values in the interval. Intuitively, the selection strategy has a significant impact on the final performance of the trained model. Here, we provide three typical strategies to select a particular value in the interval:

- Selecting the leftmost value:

$$\ell_{\text{left}}(f(\boldsymbol{x}), S) = \ell(f(\boldsymbol{x}), \underline{y}). \quad (7)$$

- Selecting the rightmost value:

$$\ell_{\text{right}}(f(\boldsymbol{x}), S) = \ell(f(\boldsymbol{x}), \overline{y}). \quad (8)$$

- Selecting the middlemost value:

$$\ell_{\text{mid}}(f(\boldsymbol{x}), S) = \ell(f(\boldsymbol{x}), \frac{\underline{y} + \overline{y}}{2}). \quad (9)$$

Obviously, these three strategies select the three most particular values (including the leftmost value, the rightmost value, and the middlemost value) in the interval. Different selection strategies could result in different errors of estimating the true target value. Given any interval example $(\boldsymbol{x}, S)$ with $|S| = a$, we analyze the mean absolute error of estimating the true target y by the three strategies when y falls at any position in the interval. We illustrate this analysis in Table 1. As shown in Table 1, since any value in the interval could be the true target, we calculate the maximum error and the expected error for each selection strategy. Clearly, if the true target is the rightmost value (i.e., $y = \overline{y}$), the rightmost selection strategy is optimal and the errors ($a/2$ and $a/4$) of the leftmost and middlemost selection strategies are maximum. If the true target is the middlemost value (i.e., $y = (\underline{y} + \overline{y})/2$), the middlemost selection strategy is optimal and both the leftmost and rightmost selection strategies achieve the error of $a/2$. If the true target is the leftmost

Table 1. Error analysis for three selection strategies (with $|S| = a$).

Strategy	Selected value	Maximum error	Expected error
Leftmost	$\underline{y}$	a	$a/2$
Rightmost	$\overline{y}$	a	$a/2$
Middlemost	$(\underline{y} + \overline{y})/2$	$a/2$	$a/4$

value (i.e., $y = \underline{y}$), the leftmost selection strategy is optimal and the errors ($a/2$ and $a/4$) of the rightmost and middlemost selection strategies are maximum. When all the values in the interval have the same probability of being the true target, the expected error of both the leftmost and rightmost selection strategies is $a/2$ and the expected error of the middlemost selection strategy is $a/4$. According to the above analysis, we can find that the middlemost selection strategy is relatively stable and can achieve a smaller error regardless of the true target value. Therefore, the middlemost selection strategy is expected to achieve better performance than the leftmost and the rightmost selection strategies, and our empirical results in Section 5 also support this argument.

Further discussion. In addition to the above middlemost selection strategy, it is natural to consider another strategy from the loss perspective, i.e., the average loss of ℓ_{left} and ℓ_{right}. Specifically, for each interval example (x, S), we can define the average loss as $\ell_{\text{avgl}}(f(\boldsymbol{x}), S) = \frac{1}{2}\ell(f(\boldsymbol{x}), \underline{y}) + \frac{1}{2}\ell(f(\boldsymbol{x}), \overline{y})$. We will theoretically analyze this method and show that only with a specific choice of the regression loss ℓ (i.e., mean absolute error), $\ell_{\text{avgl}}(f(\boldsymbol{x}), S)$ can achieve good empirical performance with theoretical guarantees.

4.2. The Statistically Consistent Limiting Method

We can find that the simple selection method only considers a single value in the interval and ignores the overall interval information. To overcome this drawback, we propose the following limiting method that limits the predicted value in the interval:

$$\ell_{\text{LM}}(f(\boldsymbol{x}), S) = \mathbb{I}\left[\underline{y} - f(\boldsymbol{x}) > 0\right] + \mathbb{I}\left[f(\boldsymbol{x}) - \overline{y} > 0\right]. \quad (10)$$

This loss function takes value 0 if $\underline{y} \leq f(\boldsymbol{x}) \leq \overline{y}$, otherwise 1. This is in line with our intention to limit the predicted values in the interval. Then, the expected regression risk of our proposed limiting method can be represented as follows:

$$R_{\text{LM}}(f) = \mathbb{E}_{\tilde{p}(\boldsymbol{x}, S)}[\ell_{\text{LM}}(f(\boldsymbol{x}), S)]. \quad (11)$$

We demonstrate that our proposed limiting method is *model-consistent*, i.e., the model learned by the limiting method from interval data converges to the optimal model learned from fully supervised data. In particular, we assume that the hypothesis space $\mathcal{F}$ is strong enough (Lv et al., 2020) such that the optimal model (i.e., $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$) in

the hypothesis space makes the optimal risk equal to 0 (i.e., $R(f^\star) = 0$). Then we introduce the following theorem.

Theorem 4.1. *Suppose that the hypothesis space $\mathcal{F}$ is strong enough (i.e., $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$ leads to $R(f^\star) = 0$). The model $f^\star_{\mathrm{LM}} = \arg\min_{f \in \mathcal{F}} R_{\mathrm{LM}}(f)$ learned by our limiting method is equivalent to the optimal model $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$.*

Theorem 4.1 demonstrates that the optimal regression model learned from fully labeled data can be identified by our limiting method given only data with interval targets (i.e., our limiting method is model consistent). However, we cannot directly train a regression model by using our limiting method in Eq. (10), since the loss function in Eq. (10) is non-convex and discontinuous. To address this problem, we propose the following surrogate loss function of our limiting method:

$$\psi_{\mathrm{LM}}(f(\boldsymbol{x}), S) = \max(0, \underline{y} - f(\boldsymbol{x})) + \max(0, f(\boldsymbol{x}) - \overline{y}).$$
(12)

As can be seen from Eq. (12), this surrogate loss is convex and is an upper bound of the original loss in Eq. (10). With the surrogate loss in Eq. (12), the expected regression risk of our proposed limiting method can be represented as:

$$R^\psi_{\mathrm{LM}}(f) = \mathbb{E}_{\tilde{p}(\boldsymbol{x}, S)}[\psi_{\mathrm{LM}}(f(\boldsymbol{x}), S)].$$

Then, we demonstrate that our limiting method with the surrogate loss is still consistent, by the following theorem.

Theorem 4.2. *Suppose the hypothesis space $\mathcal{F}$ is strong enough (i.e., $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$ leads to $R(f^\star) = 0$). The model $f^{\psi\star}_{\mathrm{LM}} = \arg\min_{f \in \mathcal{F}} R^\psi_{\mathrm{LM}}(f)$ learned by the surrogate method is equivalent to the optimal model $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$.*

Theorem 4.2 shows that the model learned by our limiting method is also equivalent to the optimal model $f^\star$ (learned from fully labeled data). This indicates that using the surrogate loss in Eq. (12), our limiting method is still consistent. Therefore, we can learn an effective regression model from the given dataset $\{\boldsymbol{x}_i, S_i\}_{i=1}^n$ by directly minimizing the following empirical risk:

$$\widehat{R}^\psi_{\mathrm{LM}}(f) = \sum_{i=1}^n \psi_{\mathrm{LM}}(f(\boldsymbol{x}), S).$$
(13)

Here, we further relate our limiting method ψ_{LM} to the average method ℓ_{avgl} discussed in Section 4.1, by the following corollary.

Corollary 4.3. *The same minimizer (model) can be derived from ψ_{LM} and ℓ_{avgl} if the mean absolute error is used as the regression loss ℓ in ℓ_{avgl}.*

Corollary 4.3 implies that with the mean absolute error, the average loss ℓ_{avgl} is also model-consistent, because our limiting method is model-consistent. However, using other losses (e.g., the mean squared error) cannot make ℓ_{avgl} theoretically grounded, and thus the empirical performance is guaranteed. We conduct experiments to demonstrate this argument, and experimental results (given in Appendix F.5) show that the mean absolute error clearly outperforms the mean squared error, when used in ℓ_{avgl}.

Consistency analysis. Here, we provide a consistency analysis for the above limiting method, which shows that the model $\widehat{f}^\psi_{\mathrm{LM}} = \arg\min_{f \in \mathcal{F}} \widehat{R}^\psi_{\mathrm{LM}}(f)$ (empirically learned from RIT data by using our limiting method) is infinite-sample consistent to the optimal model $f^\star$.

Theorem 4.4. *Assume that for all $(\boldsymbol{x}, S)$ with $S = [\underline{y}, \overline{y}]$ drawn from $\tilde{p}(\boldsymbol{x}, S)$ and all $f \in \mathcal{F}$, there exist constants M and M' such that $\max(\underline{y} - f(\boldsymbol{x}), 0) \leq M$ and $\max(f(\boldsymbol{x}) - \overline{y}, 0) \leq M'$. Suppose that the pseudo-dimensions of $\{(\boldsymbol{x}, \underline{y}) \mapsto \max(\underline{y} - f(\boldsymbol{x}), 0) \mid f \in \mathcal{F}\}$ and $\{(\boldsymbol{x}, \overline{y}) \mapsto \max(f(\boldsymbol{x}) - \overline{y}, 0) \mid f \in \mathcal{F}\}$ are finite, which are denoted by d and d'. Then, with probability at least $1 - \delta$,*

$$R^\psi_{\mathrm{LM}}(\widehat{f}^\psi_{\mathrm{LM}}) - R^\psi_{\mathrm{LM}}(f^\star) \leq 2M\sqrt{\frac{2d \log \frac{en}{d}}{n}}$$
$$+ 2M'\sqrt{\frac{2d' \log \frac{en}{d'}}{n}} + 2(M + M')\sqrt{\frac{\log \frac{4}{\delta}}{2n}}.$$

Theorem 4.4 shows that the risk of $\widehat{f}^\psi_{\mathrm{LM}}$ converges to the risk of $f^\star$, as the number of training data goes to infinity.

5. Experiments

In this section, we conduct extensive experiments to validate the effectiveness of our proposed limiting method.

5.1. Experimental Setup

Datasets. We conduct experiments on nine datasets, including two computer vision datasets (AgeDB (Moschoglou et al., 2017) and IMDB-WIKI (Rothe et al., 2018)), one natural language processing dataset (STS-B (Cer et al., 2017)), and six datasets from the UCI Machine Learning Repository (Dua & Graff, 2017) (Abalone, Airfoil, Auto-mpg, Housing, Concrete, and Power-plant). Following the data distribution proposed in Section 3.1, We generated the following RIT datasets, including AgeDB-Interval at $q = 10, 20, 30, 40$, and 50, IMDB-WIKI-Interval at $q = 20, 30$, and 40, and STS-B-Interval at $q = 3.0, 3.5, 4.0, 4.5$ and 5.0. For each UCI dataset, we selected two large values of q to generate RIT data based on the span of the label space. The specific descriptions of used datasets with the corresponding base models and the specific hyperparameter settings are reported in Appendix E.1.

Base models. For the UCI dataset, we used two models, a linear model and a multilayer perceptron (MLP), where the

Table 2. Test performance (mean and std) of each method on AgeDB. The used evaluation metrics include MSE and MAE. We repeat the sampling-and-training process 3 times. The best performance is highlighted in bold.

Metric		MSE					MAE				
Approach		$q=30$	$q=40$	$q=50$	$q=60$	$q=70$	$q=30$	$q=40$	$q=50$	$q=60$	$q=70$
Leftmost	MAE	158.38	210.34	205.75	283.83	355.59	9.88	11.59	11.27	13.62	14.93
		(6.71)	(19.85)	(22.04)	(15.36)	(105.16)	(0.22)	(0.61)	(0.80)	(0.62)	(2.38)
	MSE	134.99	196.83	221.54	295.49	347.27	9.25	11.19	11.91	13.90	14.98
		(4.03)	(19.28)	(18.04)	(32.09)	(80.77)	(0.06)	(0.71)	(0.48)	(1.24)	(2.10)
	Huber	156.34	175.95	208.03	317.17	360.11	9.85	10.54	11.42	14.39	15.41
		(17.95)	(7.61)	(18.36)	(22.57)	(53.12)	(0.57)	(0.19)	(0.61)	(0.37)	(1.30)
Rightmost	MAE	154.87	196.95	233.18	255.24	428.05	9.91	11.31	12.48	13.12	17.51
		(13.12)	(20.96)	(26.58)	(15.12)	(41.79)	(0.45)	(0.58)	(0.79)	(0.40)	(0.93)
	MSE	146.85	215.06	260.37	304.71	452.61	9.57	11.96	13.27	14.47	18.15
		(24.39)	(23.92)	(15.32)	(49.88)	(45.87)	(0.86)	(0.66)	(0.49)	(1.29)	(1.15)
	Huber	149.14	179.72	246.29	279.70	436.29	9.72	10.84	12.88	13.90	17.50
		(7.74)	(10.57)	(16.02)	(17.45)	(78.90)	(0.31)	(0.37)	(0.48)	(0.44)	(1.86)
Middlemost	MAE	116.14	133.44	129.55	138.95	150.88	8.38	8.93	8.97	9.21	9.57
		(2.57)	(5.05)	(1.37)	(5.22)	(3.66)	(0.13)	(0.15)	(0.13)	(0.11)	(0.12)
	MSE	119.90	133.27	128.84	138.36	149.82	8.45	8.94	8.89	9.32	9.53
		(6.23)	(5.18)	(3.01)	(6.10)	(5.28)	(0.18)	(0.22)	(0.21)	(0.27)	(0.07)
	Huber	121.78	131.43	131.38	140.40	149.25	8.62	8.92	8.96	9.28	9.65
		(4.75)	(3.84)	(2.20)	(6.18)	(0.70)	(0.14)	(0.15)	(0.08)	(0.19)	(0.09)
CRM		221.66	303.52	398.50	523.53	653.81	12.18	14.57	17.17	20.11	22.89
		(1.45)	(11.12)	(14.80)	(4.63)	(17.30)	(0.07)	(0.30)	(0.40)	(0.08)	(0.09)
RANN		125.04	126.02	129.86	139.83	148.25	8.69	8.73	8.89	9.32	9.69
		(1.09)	(1.74)	(1.00)	(3.41)	(2.33)	(0.04)	(0.07)	(0.05)	(0.22)	(0.11)
SINN		218.16	302.06	404.80	524.74	649.27	12.11	14.54	17.41	19.97	22.70
		(1.53)	(6.53)	(2.70)	(7.17)	(2.94)	(0.08)	(0.24)	(0.04)	(0.13)	(0.14)
IN		118.85	133.86	138.58	147.95	152.34	8.46	9.00	9.15	9.53	9.65
		(5.20)	(6.46)	(4.06)	(2.82)	(9.30)	(0.17)	(0.13)	(0.09)	(0.07)	(0.21)
LM		**115.76**	**121.24**	**123.51**	**128.04**	**129.47**	**8.36**	**8.67**	**8.75**	**8.82**	**8.92**
		(1.64)	**(4.73)**	**(2.78)**	**(3.52)**	**(7.15)**	**(0.07)**	**(0.03)**	**(0.14)**	**(0.16)**	**(0.14)**

MLP model is a five-layer (d-20-30-10-1) neural network with a ReLU activation function. For the linear model and the MLP model, we use the Adam optimization method (Kingma & Ba, 2015) with the batch size set to 512 and the number of training epochs set to 1,000, and the learning rate for all methods is selected from $\{10^{-2}, 10^{-3}\}$. For both the IMDB-WIKI and AgeDB datasets, we use ResNet-50 (He et al., 2016) as our backbone network. We use the Adam optimizer to train all methods for 100 epochs with an initial learning rate of 10^{-3} and fix the batch size to 256. For the STS-B dataset, we follow Wang et al. (2019a) to use the same 300D GloVe word embeddings and a two-layer 1500D (per direction) BiLSTM with max pooling to encode the paired sentences into independent vectors u and v, and then pass $[u; v; |u - v|; uv]$ to a regressor. We also use the Adam optimizer to train all methods for 100 epochs with an initial learning rate of 10^{-4} and fix the batch size to 256.

Compared methods. We use the leftmost, rightmost, and middlemost selection strategies analyzed in Section 4.1 as our baseline methods. Since the three methods do not rely on any loss function, we use the mean absolute error (MAE), the mean squared error (MSE), and the Huber loss (commonly used in regression tasks) as loss functions to form our baseline methods. For the Huber loss, the threshold value is selected from $\{1, 5\}$. In particular, we compare with multiple methods for interval-valued data prediction,

including CRM (Neto & De Carvalho, 2008), SINN (Yang & Wu, 2012), RANN (Yang et al., 2019), IN (Sadeghi et al., 2019). Since the outputs of these methods are intervals, we use the midpoint of the interval as the predicted value.

Evaluation metrics. For metrics, we use common evaluation metrics for regression, such as the MSE, MAE, and Pearson correlation. We also use another evaluation metric called Geometric Mean (Yang et al., 2021).

5.2. Experimental Performance

Experimental results. Table 2, Table 3 and Table 4 show some of the experimental results on the AgeDB, IMDB-WIKI, and UCI datasets, respectively. From the three tables, we have the following observations: 1) Our proposed LM outperforms all the compared methods. This verifies that our method has the ability to figure out the true real-valued labels. 2) As q increases, there is a tendency for the performance of all the methods to decrease. This is because as the size of the interval becomes larger, more interfering values are included in the interval and thus it will be more difficult to identify the true real-valued labels from the interval. 3) In our experimental setting, we set various values of q. The performance gap between our method and compared methods is more evident when q is large. This indicates that our method has stronger robustness. It is worth noting that q

Table 3. Test performance (mean and std) of each method on IMDB-WIKI. The used evaluation metrics include MSE and MAE. We repeat the sampling-and-training process 3 times. The best performance is highlighted in bold.

Metric		MSE					MAE				
Approach		$q=40$	$q=50$	$q=60$	$q=70$	$q=80$	$q=40$	$q=50$	$q=60$	$q=70$	$q=80$
Leftmost	MAE	270.15 (15.14)	336.73 (30.92)	402.05 (39.13)	494.84 (28.64)	581.54 (19.81)	13.06 (0.53)	14.76 (0.77)	16.28 (1.10)	18.30 (0.82)	20.39 (0.46)
	MSE	255.40 (13.45)	305.92 (7.33)	358.15 (25.61)	383.77 (16.82)	421.16 (94.72)	12.67 (0.40)	13.95 (0.60)	15.21 (0.57)	15.86 (0.36)	16.66 (2.36)
	Huber	291.25 (14.98)	320.37 (10.01)	385.99 (41.84)	512.73 (107.26)	596.39 (46.08)	13.39 (0.26)	14.24 (0.42)	15.82 (1.14)	18.57 (2.54)	20.71 (0.79)
Rightmost	MAE	197.26 (17.24)	265.60 (27.79)	328.77 (63.99)	608.51 (76.85)	679.26 (80.82)	11.15 (0.62)	13.38 (0.75)	14.74 (1.55)	20.80 (1.50)	22.82 (2.21)
	MSE	214.42 (10.40)	280.99 (12.52)	357.06 (64.59)	497.47 (43.98)	643.99 (85.80)	11.71 (0.35)	13.50 (0.35)	15.29 (1.54)	18.82 (1.05)	21.38 (1.49)
	Huber	198.97 (6.47)	243.81 (11.64)	379.34 (45.10)	536.87 (21.50)	491.39 (103.54)	11.23 (0.21)	12.82 (0.31)	16.19 (1.01)	19.63 (0.67)	18.24 (2.22)
Middlemost	MAE	140.29 (6.68)	148.93 (3.00)	152.20 (9.59)	154.79 (4.29)	157.17 (5.96)	8.96 (0.10)	9.24 (0.08)	9.47 (0.36)	9.55 (0.21)	9.76 (0.22)
	MSE	135.57 (2.77)	142.68 (2.43)	144.34 (2.60)	153.10 (9.34)	153.80 (3.57)	8.89 (0.10)	9.10 (0.06)	9.23 (0.11)	9.61 (0.30)	9.79 (0.08)
	Huber	142.21 (6.33)	148.52 (3.93)	153.97 (2.31)	152.52 (3.48)	154.77 (3.58)	9.04 (0.19)	9.22 (0.15)	9.49 (0.05)	9.49 (0.12)	9.70 (0.15)
CRM		302.50 (12.53)	380.12 (15.30)	519.7 (10.23)	602.65 (7.20)	740.99 (19.56)	14.38 (0.42)	16.81 (0.28)	20.11 (0.20)	21.85 (0.16)	24.61 (0.78)
RANN		137.37 (3.09)	140.79 (2.28)	145.40 (2.20)	150.11 (2.48)	166.33 (4.02)	8.98 (0.10)	8.98 (0.12)	9.32 (0.08)	9.52 (0.07)	10.09 (0.08)
SINN		314.49 (4.08)	385.29 (8.26)	515.80 (9.95)	629.51 (11.10)	754.28 (15.20)	14.79 (0.14)	16.73 (0.19)	19.84 (0.30)	22.36 (0.23)	24.73 (0.65)
IN		147.09 (0.59)	148.71 (1.01)	152.66 (2.54)	155.73 (5.67)	156.04 (2.89)	9.25 (0.04)	9.28 (0.09)	9.51 (0.10)	9.59 (0.17)	9.79 (0.06)
LM		**133.98 (1.57)**	**134.15 (2.49)**	**141.45 (2.32)**	**148.19 (2.44)**	**146.52 (5.00)**	**8.75 (0.06)**	**8.83 (0.07)**	**9.07 (0.11)**	**9.40 (0.08)**	**9.39 (0.04)**

(a) Abalone (b) Auto-mpg (c) Airfoil (d) Power-plant
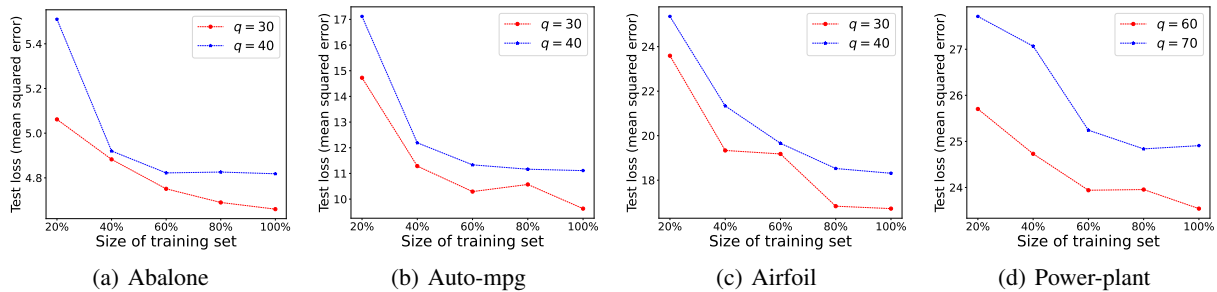
Figure 2. The test performance (MSE) on the Abalone, Auto-mpg, Airfoil and Power-plant datasets of our proposed LM when the number of training data increases.

represents the maximum interval size allowed. In real-world scenarios, a large value of q will be a more common situation because this kind of data is easier to collect. 4) The methods for interval-valued data prediction and the methods for selecting the middlemost value of the interval as the target value have similar performance. This is because when the interval predicted by the interval-valued data prediction is accurate, the middlemost value of the interval is exactly the target value of the methods that select the middlemost value of the interval as the target value.

Performance of Increasing Training Data. We demonstrate in Theorem 4.4 that the model learned by our proposed LM can converge to the optimal model learned from the fully labeled data when the number of training examples for RIT approaches infinity. To empirically validate such a theoretical finding, we further conduct experiments by changing the fraction of training examples for RIT, where 100% indicates the use of all training examples to train the model. The experimental performance of LM is shown in Figure 2, where the test loss of the model generally decreases when more training examples are used to train the model. This empirical observation accords with our theoretical analysis that the learned model will be closer to the optimal model, if more training examples are provided.

Table 4. Test performance (mean and std) of each method on the six UCI datasets trained with the MLP model. The used evaluation metrics include MSE and MAE. We repeat the sampling-and-training process 5 times. The best performance is highlighted in bold.

Dataset	Metric	q	Leftmost MSE	MAE	Huber	Rightmost MSE	MAE	Huber	Middlemost MSE	MAE	Huber	CRM	RANN	SINN	IN	LM
Abalone	MSE	30	65.33 (2.09)	55.73 (1.79)	59.82 (3.07)	7.99 (0.70)	8.08 (0.71)	8.03 (0.71)	6.38 (0.44)	5.45 (0.43)	6.01 (0.44)	6.51 (0.52)	6.67 (0.48)	6.44 (0.44)	5.61 (0.40)	**4.66 (0.49)**
		40	90.12 (3.16)	85.71 (6.28)	89.27 (4.47)	8.13 (0.67)	8.23 (0.63)	8.21 (0.64)	7.77 (0.68)	7.81 (0.76)	7.79 (0.76)	7.85 (0.81)	7.74 (0.64)	7.85 (0.81)	7.84 (0.75)	**4.81 (0.42)**
	MAE	30	7.60 (0.11)	6.96 (0.13)	7.21 (0.16)	2.01 (0.08)	2.11 (0.08)	2.10 (0.08)	1.94 (0.06)	1.77 (0.08)	1.91 (0.07)	1.94 (0.07)	1.91 (0.07)	1.94 (0.07)	1.86 (0.05)	**1.50 (0.05)**
		40	9.03 (0.15)	8.80 (0.32)	9.03 (0.18)	2.02 (0.08)	2.12 (0.07)	2.11 (0.07)	1.96 (0.08)	1.95 (0.06)	1.96 (0.06)	1.99 (0.07)	2.02 (0.08)	1.99 (0.07)	1.97 (0.06)	**1.53 (0.08)**
Airfoil	MSE	30	122.58 (89.68)	89.68 (5.26)	93.90 (8.01)	105.20 (9.19)	79.28 (7.75)	85.16 (14.44)	19.24 (1.70)	18.71 (1.16)	18.24 (2.02)	17.40 (3.07)	17.30 (2.86)	17.35 (3.13)	19.20 (1.30)	**16.72 (3.42)**
		40	200.70 (13.62)	164.64 (9.28)	183.30 (23.42)	134.36 (13.66)	115.92 (5.56)	120.34 (7.57)	24.32 (1.16)	20.43 (1.87)	20.99 (1.89)	23.12 (2.30)	22.98 (1.92)	23.08 (2.38)	19.37 (2.51)	**18.31 (2.63)**
	MAE	30	9.81 (0.52)	8.50 (0.39)	8.71 (0.62)	8.00 (0.26)	7.78 (0.38)	7.98 (0.51)	3.41 (0.24)	3.30 (0.15)	3.28 (0.25)	3.26 (0.33)	3.23 (0.32)	3.26 (0.32)	3.33 (0.16)	**3.09 (0.35)**
		40	11.93 (0.31)	11.96 (0.38)	12.18 (0.60)	8.93 (0.64)	9.24 (0.43)	9.29 (0.46)	3.97 (0.12)	3.54 (0.13)	3.61 (0.20)	3.85 (0.22)	3.85 (0.22)	3.85 (0.21)	3.47 (0.26)	**3.24 (0.27)**
Auto-mpg	MSE	30	76.60 (12.72)	34.99 (12.83)	50.43 (18.98)	51.52 (20.31)	30.41 (11.25)	24.06 (7.60)	11.59 (1.42)	11.69 (1.69)	11.52 (1.17)	11.68 (1.25)	23.60 (3.86)	11.67 (1.28)	11.60 (1.23)	**9.63 (1.62)**
		40	145.11 (19.52)	93.80 (23.82)	92.64 (32.76)	119.54 (20.53)	48.03 (15.25)	61.49 (28.88)	20.14 (5.44)	18.30 (4.52)	19.25 (4.69)	21.11 (6.18)	34.57 (8.40)	21.08 (6.30)	18.23 (4.97)	**11.11 (3.03)**
	MAE	30	7.97 (0.76)	4.47 (1.03)	5.98 (1.34)	6.33 (1.63)	4.47 (0.94)	3.89 (0.78)	2.52 (0.23)	2.51 (0.18)	2.48 (0.18)	2.50 (0.19)	3.76 (0.38)	2.50 (0.19)	2.50 (0.17)	**2.19 (0.18)**
		40	11.59 (1.02)	8.09 (1.16)	8.16 (1.84)	10.91 (1.75)	5.78 (1.03)	6.68 (1.92)	3.46 (0.48)	3.15 (0.37)	3.18 (0.43)	3.39 (0.50)	4.62 (0.59)	3.36 (0.52)	3.24 (0.41)	**2.31 (0.30)**
Housing	MSE	30	55.46 (8.52)	52.66 (15.47)	52.79 (7.06)	75.37 (27.20)	78.40 (8.87)	83.83 (15.28)	27.49 (11.60)	26.24 (8.20)	25.07 (5.82)	27.58 (6.80)	25.45 (8.95)	26.32 (10.12)	25.06 (7.59)	**22.13 (3.71)**
		40	88.85 (22.73)	101.70 (13.36)	83.96 (15.88)	109.01 (32.28)	124.34 (12.16)	124.14 (12.18)	30.70 (4.04)	35.84 (2.72)	32.63 (4.86)	32.66 (4.47)	31.15 (5.46)	34.92 (4.47)	33.00 (5.71)	**24.53 (8.00)**
	MAE	30	6.02 (0.65)	5.61 (1.06)	5.49 (0.61)	6.88 (0.54)	7.63 (0.33)	8.31 (0.85)	3.52 (0.36)	3.54 (0.45)	3.50 (0.27)	3.52 (0.50)	3.51 (0.56)	3.63 (0.66)	3.56 (0.43)	**3.47 (0.57)**
		40	7.15 (0.66)	7.46 (0.53)	7.17 (0.62)	7.87 (0.92)	8.27 (0.42)	8.26 (0.42)	4.06 (0.45)	4.41 (0.28)	4.11 (0.31)	4.14 (0.45)	3.95 (0.48)	4.25 (0.39)	4.19 (0.46)	**3.49 (0.48)**
Concrete	MSE	70	283.72 (13.93)	291.78 (17.85)	296.00 (7.37)	278.35 (18.24)	278.30 (18.16)	278.26 (18.18)	70.94 (8.19)	72.84 (7.78)	73.01 (6.65)	74.68 (5.80)	74.16 (4.23)	71.68 (5.84)	67.22 (5.82)	**58.45 (3.52)**
		80	420.44 (68.96)	424.03 (57.64)	423.70 (40.60)	284.03 (15.41)	284.04 (15.41)	284.02 (15.43)	88.26 (6.35)	95.09 (8.56)	86.54 (13.11)	86.40 (3.13)	80.77 (2.78)	83.38 (3.13)	92.48 (16.38)	**59.47 (7.81)**
	MAE	70	13.93 (1.04)	14.07 (0.99)	14.42 (0.64)	13.38 (0.49)	13.38 (0.49)	13.38 (0.49)	6.70 (0.54)	6.81 (0.25)	6.83 (0.32)	6.78 (0.81)	6.51 (0.82)	6.77 (0.82)	6.38 (0.34)	**5.86 (0.41)**
		80	17.74 (1.55)	17.67 (1.48)	17.93 (1.26)	13.47 (0.38)	13.47 (0.38)	13.47 (0.38)	7.42 (0.46)	7.62 (0.33)	7.35 (0.61)	7.40 (0.68)	7.26 (0.33)	7.48 (0.47)	7.49 (0.70)	**5.71 (0.34)**
Power-plant	MSE	60	264.98 (49.48)	201.00 (51.72)	245.19 (35.67)	262.00 (107.86)	246.15 (69.44)	268.46 (53.87)	31.96 (2.69)	28.95 (3.03)	29.38 (3.76)	32.85 (3.78)	31.30 (2.84)	32.66 (3.95)	29.36 (3.41)	**23.54 (1.41)**
		70	279.48 (3.14)	316.33 (72.10)	337.53 (61.63)	280.29 (76.53)	269.03 (81.76)	293.51 (2.20)	48.82 (1.53)	39.41 (2.76)	40.83 (3.41)	48.53 (1.20)	47.62 (2.37)	48.62 (1.22)	40.04 (3.13)	**24.91 (0.60)**
	MAE	60	14.48 (0.10)	12.55 (1.58)	14.08 (0.34)	13.92 (3.31)	13.98 (2.39)	14.60 (1.43)	4.55 (0.18)	4.30 (0.21)	4.36 (0.32)	4.59 (0.28)	4.49 (0.21)	4.58 (0.31)	4.37 (0.29)	**3.84 (0.11)**
		70	14.29 (0.16)	14.48 (1.51)	14.70 (0.94)	14.17 (1.85)	14.70 (2.56)	14.99 (0.10)	5.59 (0.14)	4.94 (0.18)	5.13 (0.20)	5.62 (0.08)	5.51 (0.10)	5.62 (0.07)	5.01 (0.16)	**3.96 (0.05)**

More experimental results. We provide more experimental results including the comparison results with the fully supervised method, maximum margin interval trees method (MMIT (Drouin et al., 2017)), and more results on evaluation metrics and models in Appendix F. These results also demonstrate the effectiveness of our method.

6. Conclusion

In this paper, we studied an interesting weakly supervised regression setting called *regression with interval targets* (RIT). For the RIT setting, we first proposed a novel statistical model to describe the data generation process for RIT and demonstrated its validity. The explicitly derived data distribution can be helpful to empirical risk minimization. Then, we analyzed a simple selection method that selects a particular value in the interval as the target value to train the model. We empirically showed that this simple method could work well if the middlemost value in the interval is selected. Afterward, we proposed a statistically consistent limiting method to train the model by limiting the predictions to the interval. We further derived an estimation error bound for this method. Finally, we conducted extensive experiments on various datasets to demonstrate the effectiveness of our proposed method. In future work, it would be interesting to study a harder setting of RIT, where the true target value might be outside the given interval.

Acknowledgements

References

Amar, R. A., Dooly, D. R., Goldman, S. A., and Zhang, Q. Multiple-instance learning of real-valued data. In *ICML*, pp. 3–10, 2001.

Billard, L. Symbolic data analysis: what is it? In *Compstat 2006-Proceedings in Computational Statistics: 17th Symposium Held in Rome, Italy, 2006*, pp. 261–269. Springer, 2006.

Billard, L. and Diday, E. Regression analysis for interval-valued data. In *Data Analysis, Classification, and Related Methods*, pp. 369–374. Springer, 2000.

Bock, H.-H. and Diday, E. *Analysis of symbolic data: exploratory methods for extracting statistical information from complex data*. Springer Science & Business Media, 1999.

Carpentier, A. and Schlüter, T. Learning relationships between data obtained independently. In *AISTATS*, pp. 658–666, 2016.

Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., and Specia, L. Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation. *arXiv preprint arXiv:1708.00055*, 2017.

Cour, T., Sapp, B., and Taskar, B. Learning from partial labels. *The Journal of Machine Learning Research*, 12: 1501–1536, 2011.

Drouin, A., Hocking, T., and Laviolette, F. Maximum margin interval trees. In *NeurIPS*, 2017.

Dua, D. and Graff, C. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

Fagundes, R. A., De Souza, R. M., and Cysneiros, F. J. A. Interval kernel regression. *Neurocomputing*, 128:371–388, 2014.

Feng, L., Lv, J., Han, B., Xu, M., Niu, G., Geng, X., An, B., and Sugiyama, M. Provably consistent partial-label learning. In *NeurIPS*, pp. 10948–10960, 2020.

Geng, X., Yin, C., and Zhou, Z.-H. Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10): 2401–2412, 2013.

Giordani, P. Lasso-constrained regression analysis for interval-valued data. *Advances in Data Analysis and Classification*, 9(1):5–19, 2015.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *CVPR*, pp. 770–778, 2016.

Hu, W., Li, Z., and Yu, D. Simple and effective regularization methods for training on noisily labeled data with generalization guarantee. In *ICLR*, 2020.

Ishibuchi, H. and Tanaka, H. An extension of the bp-algorithm to interval input vectors-learning from numerical data and expert's knowledge. In *IJCNN*, pp. 1588–1593, 1991.

Ishida, T., Niu, G., Menon, A., and Sugiyama, M. Complementary-label learning for arbitrary losses and models. In *ICML*, pp. 2971–2980, 2019.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *ICLR*, 2015.

Kleinbaum, D. G., Klein, M., et al. *Survival analysis: a self-learning text*, volume 3. Springer, 2012.

Kostopoulos, G., Karlos, S., Kotsiantis, S., and Ragos, O. Semi-supervised regression: A recent review. *Journal of Intelligent & Fuzzy Systems*, 35(2):1483–1500, 2018.

Lauro, C. N. and Palumbo, F. Principal component analysis of interval data: a symbolic data analysis approach. *Computational Statistics*, 15(1):73–87, 2000.

Lesaffre, E., Komárek, A., and Declerck, D. An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Statistical Methods in Medical Research*, 14(6):539–552, 2005.

Li, Y.-F., Zha, H.-W., and Zhou, Z.-H. Learning safe prediction for semi-supervised regression. In *AAAI*, 2017.

Lindsey, J. C. and Ryan, L. M. Methods for interval-censored data. *Statistics in Medicine*, 17(2):219–238, 1998.

Lv, J., Xu, M., Feng, L., Niu, G., Geng, X., and Sugiyama, M. Progressive identification of true labels for partial-label learning. In *ICML*, 2020.

Machin, D., Cheung, Y. B., and Parmar, M. *Survival analysis: a practical approach*. John Wiley & Sons, 2006.

Manski, C. F. and Tamer, E. Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2):519–546, 2002.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. *Foundations of Machine Learning*. MIT Press, 2012.

Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., and Zafeiriou, S. Agedb: The first manually collected, in-the-wild age database. In *CVPRW*, pp. 1997–2005, 2017.

Neto, E. d. A. L. and De Carvalho, F. D. A. Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis*, 52(3):1500–1515, 2008.

Neto, E. d. A. L. and De Carvalho, F. D. A. Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics & Data Analysis*, 54 (2):333–347, 2010.

Park, S., Wang, X., Lim, J., Xiao, G., Lu, T., and Wang, T. Bayesian multiple instance regression for modeling immunogenic neoantigens. *Statistical Methods in Medical Research*, 29(10):3032–3047, 2020.

Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., and Qu, L. Making deep neural networks robust to label noise: A loss correction approach. In *CVPR*, pp. 1944–1952, 2017.

Rabinowitz, D., Tsiatis, A., and Aragon, J. Regression with interval-censored data. *Biometrika*, 82(3):501–513, 1995.

Ristovski, K., Das, D., Ouzienko, V., Guo, Y., and Obradovic, Z. Regression learning with multiple noisy oracles. In *ECAI*, pp. 445–450, 2010.

Rothe, R., Timofte, R., and Van Gool, L. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2):144–157, 2018.

Sadeghi, J., De Angelis, M., and Patelli, E. Efficient training of interval neural networks for imprecise training data. *Neural Networks*, 118:338–351, 2019.

Stulp, F. and Sigaud, O. Many regression algorithms, one unified model: A review. *Neural Networks*, 69:60–79, 2015.

Sun, J. *The statistical analysis of interval-censored failure time data*. Springer, 2006.

Uysal, I. and Güvenir, H. A. An overview of regression techniques for knowledge discovery. *The Knowledge Engineering Review*, 14(4):319–340, 1999.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *ICLR*, 2019a.

Wang, P., Li, Y., and Reddy, C. K. Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 51 (6):1–36, 2019b.

Wang, Z., Lan, L., and Vucetic, S. Mixture model for multiple instance regression and applications in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6):2226–2237, 2011.

Wasserman, L. and Lafferty, J. Statistical analysis of semi-supervised regression. In *NeurIPS*, 2007.

Xu, L., Honda, J., Niu, G., and Sugiyama, M. Uncoupled regression from pairwise comparison data. In *NeurIPS*, 2019.

Yang, D. and Liu, Y. L1/2 regularization learning for smoothing interval neural networks: Algorithms and convergence analysis. *Neurocomputing*, 272:122–129, 2018.

Yang, D. and Wu, W. A smoothing interval neural network. *Discrete Dynamics in Nature and Society*, 2012, 2012.

Yang, Y., Zha, K., Chen, Y., Wang, H., and Katabi, D. Delving into deep imbalanced regression. In *ICML*, pp. 11842–11851, 2021.

Yang, Z., Lin, D. K., and Zhang, A. Interval-valued data prediction via regularized artificial neural network. *Neurocomputing*, 331:336–345, 2019.

A. Proofs about the Problem Setting

A.1. Prove of Theorem 3.1

For a specific label y, we define the set of all the possible intervals whose size is less than q as

$$\mathcal{S}_q^y = \{S \mid S \in \mathcal{S}, |S| \le q, y \in S\}$$

Since $\mathcal{S}_q^y$ is a continuous space, we use the sum of the number of all possible intervals to represent the size of $\mathcal{S}_q^y$, that is, the integral over all possible intervals. We might as well discuss the range of values of $\underline{y}$ and then fix $\underline{y}$ to discuss the values of $\overline{y}$. We can easily know that $\underline{y} \in [y - q, y]$. If $\underline{y} < y - q$, even the largest interval $S' = [\underline{y}, \underline{y} + q]$ cannot contain y ($\underline{y} + q < y$). If $\underline{y} > y$, then the interval must not contain y ($\underline{y} > y$). After determining $\underline{y}$, the maximum value of $\overline{y}$ is $\underline{y} + q$ ($\overline{y} - \underline{y} \le q$) and the minimum value is y ($y \le \overline{y}$), so $\overline{y} \in [y, \underline{y} + q]$, then $|\mathcal{S}_q^y| = \int_{y-q}^y \int_y^{\underline{y}+q} 1 \, \mathrm{d}\overline{y}\mathrm{d}\underline{y} = \frac{q^2}{2}$. From our formulation of the interval data distribution $\tilde{p}(\boldsymbol{x}, S)$, we can obtain the simplified expression $\tilde{p}(\boldsymbol{x}, S) = \frac{2}{q^2} \int_{y \in S} p(\boldsymbol{x}, y) \, \mathrm{d}y$. Then, we have

$$
\begin{aligned}
\int_{\mathcal{S}} \int_{\mathcal{X}} \tilde{p}(\boldsymbol{x}, S)\mathrm{d}\boldsymbol{x} \, \mathrm{d}S &= \int_{\mathcal{X}} \int_{\mathcal{S}} \int_{y \in S} \frac{2}{q^2} p(\boldsymbol{x}, y) \, \mathrm{d}y\mathrm{d}S\mathrm{d}\boldsymbol{x} \\
&= \frac{2}{q^2} \int_{\mathcal{X}} \int_{\mathcal{Y}} \int_{\mathcal{S}_q^y} p(\boldsymbol{x}, y) \, \mathrm{d}S\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= \frac{2}{q^2} \int_{\mathcal{X}} \int_{\mathcal{Y}} p(\boldsymbol{x}, y) \int_{\mathcal{S}_q^y} 1 \, \mathrm{d}S\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= \frac{2}{q^2} \int_{\mathcal{X}} \int_{\mathcal{Y}} p(\boldsymbol{x}, y)|\mathcal{S}_q^y|\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= \frac{2}{q^2} \int_{\mathcal{X}} \int_{\mathcal{Y}} p(\boldsymbol{x}, y)\frac{q^2}{2}\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} p(\boldsymbol{x}, y)\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= 1,
\end{aligned}
$$

which concludes the proof of Theorem 3.1. $\qquad\square$

A.2. Prove of Theorem 3.2

It is intuitive to express $\Pr[y \in S|\boldsymbol{x}, S]$ as

$$
\begin{aligned}
\Pr[y \in S|\boldsymbol{x}, S] &= 1 - \Pr[y \notin S|\boldsymbol{x}, S] \\
&= 1 - \int_{y \notin S} p(y|\boldsymbol{x}, S)\mathrm{d}y \\
&= 1 - \int_{y \notin S} \frac{p(S|y, \boldsymbol{x})p(y, \boldsymbol{x})}{p(S|\boldsymbol{x})}\mathrm{d}y \\
&= 1 - \int_{y \notin S} \frac{p(S|y)p(y, \boldsymbol{x})}{\int_{y' \in S} p(S|y')p(y'|\boldsymbol{x})\mathrm{d}y'}\mathrm{d}y \\
&= 1 - |\mathcal{S}_q^{y'}| \int_{y \notin S} \frac{p(S|y)p(y, \boldsymbol{x})}{\int_{y \in S} p(y'|\boldsymbol{x})\mathrm{d}y'}\mathrm{d}y \\
&= 1 - \frac{q^2}{2} \int_{y \notin S} \frac{p(S|y)p(y, \boldsymbol{x})}{\int_{y' \in S} p(y'|\boldsymbol{x})\mathrm{d}y'}\mathrm{d}y \\
&= 1,
\end{aligned}
$$

where the last equality holds because $p(S|y) = 0$ if $y \notin S$, in terms of Eq. (3). Which concludes the proof of Theorem 3.2. $\qquad\square$

A.3. Prove of Lemma 3.3

We consider the case where the true label y is a specific value, then we have

$$
\begin{aligned}
p(y \in S, y|\boldsymbol{x}) &= \Pr[y \in S|\boldsymbol{x}, y]p(y|\boldsymbol{x}) \\
&= \int_S p(y \in S, S|\boldsymbol{x}, y)p(y|\boldsymbol{x}) \, \mathrm{d}S \\
&= \int_S \Pr[y \in S|\boldsymbol{x}, y, S]p(y|\boldsymbol{x})p(S|\boldsymbol{x}, y) \, \mathrm{d}S \\
&= \int_S \Pr[y \in S|\boldsymbol{x}, y, S]p(y|\boldsymbol{x})p(S) \, \mathrm{d}S
\end{aligned}
$$

where the last equality holds due to the fact that for each example $(\boldsymbol{x}, y)$, S is uniformly and randomly chosen, if q is specific, $p(S) = \frac{1}{|S_q|}$. As with $\mathcal{S}_q^y$, we integrate over all possible intervals to calculate the size of $|\mathcal{S}_q|$. We can easily know that when $\underline{y} \in [y_{\min} - q, y_{\min}]$, $\overline{y} \in [y_{\min}, \underline{y} + q]$, when $\underline{y} \in [y_{\min}, y_{\max}]$, $\overline{y} \in [\underline{y}, \underline{y} + q]$, so $|S_q| = \int_{y_{\min}-q}^{y_{\min}} \int_{y_{\min}}^{\underline{y}+q} 1 \, \mathrm{d}y_r \mathrm{d}\underline{y} + \int_{y_{\min}}^{y_{\max}} \int_{\underline{y}}^{\underline{y}+q} 1 \, \mathrm{d}\overline{y}\mathrm{d}\underline{y} = \frac{1}{2}q^2 + q(y_{\max} - y_{\min})$, we have

$$
\begin{aligned}
p(y \in S, y|\boldsymbol{x}) &= \int_S \Pr[y \in S|\boldsymbol{x}, y, S]p(y|\boldsymbol{x})p(S) \, \mathrm{d}S \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} \int_S \Pr[y \in S|\boldsymbol{x}, y, S]p(y|\boldsymbol{x}) \, \mathrm{d}S \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} \int_S \Pr[y \in S|\boldsymbol{x}, y, S]\mathrm{d}S \, p(y|\boldsymbol{x}) \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} |\mathcal{S}_q^y|p(y|\boldsymbol{x}) \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} \frac{q^2}{2}p(y|\boldsymbol{x}) \quad (\because |\mathcal{S}_q^y| = \frac{q^2}{2}) \\
&= \frac{q}{2(y_{\max} - y_{\min}) + q}p(y|\boldsymbol{x})
\end{aligned}
$$

By integrating y on both sides, we can obtain

$$
\Pr[y \in S|\boldsymbol{x}] = \frac{q}{2(y_{\max} - y_{\min}) + q}
$$

which concludes the proof of Lemma 3.3. □

A.4. Prove of Theorem 3.4

Let us express $p(S|y \in S, \boldsymbol{x})$ as

$$
\begin{aligned}
p(S|y \in S, \boldsymbol{x}) &= \frac{p(y \in S, S|\boldsymbol{x})}{\Pr[y \in S|\boldsymbol{x}]} \\
&= \frac{\Pr[y \in S|S, \boldsymbol{x}]p(S|\boldsymbol{x})}{\Pr[y \in S|\boldsymbol{x}]} \\
&= \frac{\Pr[y \in S|S, \boldsymbol{x}]p(S)}{\Pr[y \in S|\boldsymbol{x}]}
\end{aligned}
$$

where the last equality holds due to the fact that for each instance $\boldsymbol{x}$, S is uniformly and randomly chosen. Since $p(S) = \frac{1}{|S_q|}$ if q is specific. We can easily know that when $\underline{y} \in [y_{\min} - q, y_{\min}], \overline{y} \in [y_{\min}, \underline{y} + q]$, when $\underline{y} \in [y_{\min}, y_{\max}], \overline{y} \in [\underline{y}, \underline{y} + q]$, so $|S_q| = \int_{y_{\min}-q}^{y_{\min}} \int_{y_{\min}}^{\underline{y}+q} 1 \, \mathrm{d}\overline{y}\mathrm{d}\underline{y} + \int_{y_{\min}}^{y_{\max}} \int_{\underline{y}}^{\underline{y}+q} 1 \, \mathrm{d}\overline{y}\mathrm{d}\underline{y} = \frac{1}{2}q^2 + q(y_{\max} - y_{\min})$, we have

$$
\begin{aligned}
p(S|y \in S, \boldsymbol{x}) &= \frac{\Pr[y \in S|S, \boldsymbol{x}]p(S)}{\Pr[y \in S|\boldsymbol{x}]} \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} \frac{\Pr[y \in S|S, \boldsymbol{x}]}{\Pr[y \in S|\boldsymbol{x}]} \\
&= \frac{2}{2q(y_{\max} - y_{\min}) + q^2} \frac{2(y_{\max} - y_{\min} + q)}{q} \Pr[y \in S|S, \boldsymbol{x}] \quad (by\ Lemma\ 3.3) \\
&= \frac{2}{q^2} \Pr[y \in S|S, \boldsymbol{x}] \\
&= \int_{y \in S} \frac{2}{q^2} p(y|\boldsymbol{x}) \mathrm{d}y
\end{aligned}
$$

By multiplying $p(\boldsymbol{x})$ on both side, we have

$$
\begin{aligned}
p(\boldsymbol{x}, S|y \in S) &= \int_{y \in S} \frac{2}{q^2} p(\boldsymbol{x}, y) \mathrm{d}y \\
&= \int_{y_l}^{y_r} \frac{2}{q^2} p(\boldsymbol{x}, y) \mathrm{d}y \\
&= \tilde{p}(\boldsymbol{x}, S)
\end{aligned}
$$

which concludes the proof of Theorem 3.4. $\qquad\square$

B. Proofs of The Model Consistent

B.1. Prove of Theorem 4.1

First, we prove that the optimal model $f^\star$ learned from ordinary regression expected risk (1) is also the optimal model for $R_{LM}(f)$ as follows.

$$
\begin{aligned}
R_{\mathrm{LM}}(f^\star) &= \mathbb{E}_{\tilde{p}(\boldsymbol{x}, S)}[\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S)] \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \tilde{p}(\boldsymbol{x}, S)\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S) \mathrm{d}S \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \int_{\mathcal{Y}} p(\boldsymbol{x}, y, S)\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S) \mathrm{d}y \mathrm{d}S \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \int_{\mathcal{Y}} p(S|\boldsymbol{x}, y)p(y|\boldsymbol{x})p(\boldsymbol{x})\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S) \mathrm{d}y \mathrm{d}S \mathrm{d}\boldsymbol{x} \quad (14) \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} p(y|\boldsymbol{x})p(\boldsymbol{x})\ell(f^\star(\boldsymbol{x}), y) \mathrm{d}y \mathrm{d}\boldsymbol{x} \int_{\mathcal{S}} p(S|\boldsymbol{x}, y) \mathrm{d}S \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} \ell(f^\star(\boldsymbol{x}), y)p(\boldsymbol{x}, y) \mathrm{d}y \mathrm{d}\boldsymbol{x} \\
&= R(f^\star) = 0
\end{aligned}
$$

where we used the equality $\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S) = \ell(f^\star(\boldsymbol{x}), y)$. This because when the true label $y \in S$, $\ell_{\mathrm{LM}}(f^\star(\boldsymbol{x}), S) = \mathbb{I}_{\{\underline{y} - f^\star(\boldsymbol{x}) > 0\}} + \mathbb{I}_{\{f^\star(\boldsymbol{x}) - \overline{y} > 0\}} = \ell(f^\star(\boldsymbol{x}), y) = 0$. Therefore $f^\star$ is the optimal model for R_{LM}.

On the other hand, we prove that $f^\star$ is the sole optimal model for R_{LM} by contradiction. Specifically, we assume that there is at least one other model g that makes $R_{\mathrm{LM}}(g) = 0$ and predicts a label $y_g \neq y$ for at least one instance $\boldsymbol{x}$. Therefore, for any S containing y_g we have

$$
\ell_{\mathrm{LM}}(g(\boldsymbol{x}), S) = \mathbb{I}_{\{\underline{y} - y_g > 0\}} + \mathbb{I}_{\{y_g - \overline{y} > 0\}} = 0 \quad (15)
$$

Nevertheless, the above equality could be always true on the condition that y_g is invariably included in the interval S of $\boldsymbol{x}$. In the problem setting, there is no other false label that *always* occurs with true label in the interval S. Therefore, there is one, and only one minimizer of R_{LM}, which is the same as the minimizer $f^\star$ learned from fully labeled data. The proof is completed. $\qquad\square$

B.2. Prove of Theorem 4.2

First, we prove that the optimal model $f_{\mathrm{LM}}^\star$ learned from limiting method expected risk (11) is also the optimal model for $R_{\mathrm{LM}}^\psi(f)$ as follows.

$$
\begin{aligned}
R_{\mathrm{LM}}^\psi(f_{\mathrm{LM}}^\star) &= \mathbb{E}_{\tilde{p}(\boldsymbol{x},S)}[\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S)] \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \tilde{p}(\boldsymbol{x},S)\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S)\mathrm{d}S\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \int_{\mathcal{Y}} p(\boldsymbol{x},y,S)\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S)\mathrm{d}y\mathrm{d}S\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{S}} \int_{\mathcal{Y}} p(S|\boldsymbol{x},y)p(y|\boldsymbol{x})p(\boldsymbol{x})\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S)\mathrm{d}y\mathrm{d}S\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} p(y|\boldsymbol{x})p(\boldsymbol{x})\ell_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),y)\mathrm{d}y\mathrm{d}\boldsymbol{x} \int_{\mathcal{S}} p(S|\boldsymbol{x},y)\mathrm{d}S \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} \ell_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),y)p(\boldsymbol{x},y)\mathrm{d}y\mathrm{d}\boldsymbol{x} \\
&= R_{\mathrm{LM}}(f_{\mathrm{LM}}^\star) = 0
\end{aligned}
\tag{16}
$$

where we used the equality $\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S) = \ell_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),y)$. This because when the true label $y \in S$, $\psi_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),S) = \max(0,\underline{y} - f_{\mathrm{LM}}^\star(\boldsymbol{x})) + \max(0,f_{\mathrm{LM}}^\star(\boldsymbol{x}) - \overline{y}) = \ell_{\mathrm{LM}}(f_{\mathrm{LM}}^\star(\boldsymbol{x}),y) = 0$. Therefore $f_{\mathrm{LM}}^\star$ is the optimal model for R_{LM}^ψ.

On the other hand, we prove that $f_{\mathrm{LM}}^\star$ is the sole optimal model for R_{LM}^ψ by contradiction. Specifically, we assume that there is at least one other model h that makes $R_{\mathrm{LM}}^\psi(h) = 0$ and predicts a label $y_h \neq y$ for at least one instance $\boldsymbol{x}$. Therefore, for any S containing y_h we have

$$
\psi_{\mathrm{LM}}(h(\boldsymbol{x}),S) = \max(0,\underline{y} - y_h) + \max(0,y_h - \overline{y}) = 0
\tag{17}
$$

Nevertheless, the above equality could be always true on the condition that y_h is invariably included in the interval S of $\boldsymbol{x}$. In the problem setting, there is no other false label that *always* occurs with true label in the interval S. Therefore, there is one, and only one minimizer of R_{LM}^ψ, which is the same as the minimizer $f_{\mathrm{LM}}^\star$ learned from limiting method. By Theorem 4.1, $f_{\mathrm{LM}}^\star$ is the same as the minimizer $f^\star$ learned from fully labeled data. The proof is completed. $\qquad\square$

C. Proof of Corollary 4.3

For any interval instance $(\boldsymbol{x},S)$, We consider three possible cases of model prediction: the predicted value is on the left side of the interval ($f(x) < \underline{y}$), the predicted value is on the right side of the interval ($f(x) > \overline{y}$) and the predicted value is exactly inside the interval ($\underline{y} \leq f(x) \leq \overline{y}$).

If the predicted value of the model lie on the left side of the interval, the losses of AVGL_MAE and the surrogate method are as follows.

$$
\ell_{\mathrm{avgl\_mae}}(f(\boldsymbol{x}),S) = \frac{1}{2}(|f(\boldsymbol{x}) - \underline{y}| + |f(\boldsymbol{x}) - \overline{y}|) = \frac{1}{2}(\underline{y} - f(\boldsymbol{x}) + \overline{y} - f(\boldsymbol{x})) = \frac{\underline{y} + \overline{y}}{2} - f(\boldsymbol{x}),
$$
$$
\psi_{\mathrm{LM}}(f(\boldsymbol{x}),S) = \max(0,\underline{y} - f(\boldsymbol{x})) + \max(0,f(\boldsymbol{x}) - \overline{y}) = \underline{y} - f(\boldsymbol{x}) + 0 = \underline{y} - f(\boldsymbol{x}).
$$

If the predicted value of the model lie on the right side of the interval, the losses of AVGL_MAE and the surrogate method

are as follows.

$$\ell_{\text{avgl\_mae}}(f(\boldsymbol{x}), S) = \frac{1}{2}(|f(\boldsymbol{x}) - \underline{y}| + |f(\boldsymbol{x}) - \overline{y}|) = \frac{1}{2}(f(\boldsymbol{x}) - \underline{y} + f(\boldsymbol{x}) - \overline{y}) = f(\boldsymbol{x}) - \frac{\underline{y} + \overline{y}}{2},$$

$$\psi_{\text{LM}}(f(\boldsymbol{x}), S) = \max(0, \underline{y} - f(\boldsymbol{x})) + \max(0, f(\boldsymbol{x}) - \overline{y}) = 0 + f(\boldsymbol{x}) - \overline{y} = f(\boldsymbol{x}) - \overline{y}.$$

If the predicted value of the model lie in the interval, the losses of AVGL_MAE and the surrogate method are as follows.

$$\ell_{\text{avgl\_mae}}(f(\boldsymbol{x}), S) = \frac{1}{2}(|f(\boldsymbol{x}) - \underline{y}| + |f(\boldsymbol{x}) - \overline{y}|) = 0$$

$$\psi_{\text{LM}}(f(\boldsymbol{x}), S) = \max(0, \underline{y} - f(\boldsymbol{x})) + \max(0, f(\boldsymbol{x}) - \overline{y}) = 0$$

We can see that the losses of AVGL_MAE and the surrogate loss differ only in constant terms on the three possible cases. If our training model uses gradient descent, the gradients of AVGL_MAE and the surrogate method are the same on all three possible cases. $\qquad\square$

D. Proof of Theorem 4.4

Before directly proving Theorem 4.4, we first introduce the following lemma.

Lemma D.1. *Let $\widehat{f}$ be the empirical risk minimizer (i.e., $\widehat{f} = \arg\min_{f \in \mathcal{F}} \widehat{R}(f)$) and $f^\star$ be the true risk minimizer (i.e., $f^\star = \arg\min_{f \in \mathcal{F}} R(f)$), then the following inequality holds:*

$$R(\widehat{f}) - R(f^\star) \leq 2 \sup_{f \in \mathcal{F}} \left| \widehat{R}(f) - R(f) \right|.$$

Proof. It is intuitive to obtain

$$\begin{aligned} R(\widehat{f}) - R(f^\star) &\leq R(\widehat{f}) - \widehat{R}(\widehat{f}) + \widehat{R}(\widehat{f}) - R(f^\star) \\ &\leq R(\widehat{f}) - \widehat{R}(\widehat{f}) + R(\widehat{f}) - R(f^\star) \\ &\leq 2 \sup_{f \in \mathcal{F}} \left| \widehat{R}(f) - R(f) \right|, \end{aligned}$$

which completes the proof. The same proof has been provided in Mohri et al. (2012). $\qquad\square$

Recall that $R_{\text{LM}}^\psi(f)$ is denoted by

$$\begin{aligned} R_{\text{LM}}^\psi(f) &= \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\psi(\underline{y} - f(\boldsymbol{x})) + \psi(f(\boldsymbol{x}) - \overline{y}) \right] \\ &= \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\max(\underline{y} - f(\boldsymbol{x}), 0) + \max(f(\boldsymbol{x}) - \overline{y}, 0) \right] \\ &= \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\max(\underline{y} - f(\boldsymbol{x}), 0) \right] + \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\max(f(\boldsymbol{x}) - \overline{y}, 0) \right] \\ &= R_{\text{LM}}^{\psi, l}(f) + R_{\text{LM}}^{\psi, r}(f), \end{aligned}$$

where we have introduced $R_{\text{LM}}^{\psi, l}(f) = \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\max(\underline{y} - f(\boldsymbol{x}), 0) \right]$ and $R_{\text{LM}}^{\psi, r}(f) = \mathbb{E}_{p(\boldsymbol{x}, \underline{y}, \overline{y})} \left[\max(f(\boldsymbol{x}) - \overline{y}, 0) \right]$ in the last equality. In this way, we have

$$\begin{aligned} R_{\text{LM}}^\psi(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^\psi(f^\star) &= R_{\text{LM}}^\psi(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^\psi(f_{\text{LM}}^\star) \\ &\leq 2 \sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^\psi(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^\psi(f_{\text{LM}}^\star) \right| \\ &\leq 2 \sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi, l}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi, l}(f_{\text{LM}}^\star) \right| + 2 \sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi, r}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi, r}(f_{\text{LM}}^\star) \right| \end{aligned}$$

where the first equality holds, and the last inequality means that we can directly bound $\sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi, l}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi, l}(f_{\text{LM}}^\star) \right|$ and $\sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi, r}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi, r}(f_{\text{LM}}^\star) \right|$. Based on the assumptions introduced in Theorem 4.4 and using the discussion of

Theorem 10.6 in Mohri et al. (2012), with probability $1 - \delta$,

$$\sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi,l}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi,l}(f_{\text{LM}}^{\star}) \right| \leq M\sqrt{\frac{2d \log \frac{en}{d}}{n}} + M\sqrt{\frac{\log \frac{2}{\delta}}{2n}},$$

$$\sup_{f \in \mathcal{F}} \left| R_{\text{LM}}^{\psi,r}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi,r}(f_{\text{LM}}^{\star}) \right| \leq M'\sqrt{\frac{2d' \log \frac{en}{d'}}{n}} + M'\sqrt{\frac{\log \frac{2}{\delta}}{2n}}.$$

Therefore, with probability $1 - \delta$,

$$R_{\text{LM}}^{\psi}(\widehat{f}_{\text{LM}}) - R_{\text{LM}}^{\psi}(f^{\star}) \leq 2M\sqrt{\frac{2d \log \frac{en}{d}}{n}} + 2M'\sqrt{\frac{2d' \log \frac{en}{d'}}{n}} + 2(M + M')\sqrt{\frac{\log \frac{4}{\delta}}{2n}},$$

which completes the proof of Theorem 4.4.

E. Additional Information of Experiments

E.1. Details of Datasets

In our experiments, we used AgeDB, IMDB-WIKI, STS-B and 6 UCI benchmark datasets including Abalone, Airfoil, Auto-mpg, Housing, Concrete and Power-plant. For each dataset, we follow the data distribution proposed in Section 3.1 to generate interval data. Then we randomly split each dataset into training, validation, and test sets by the proportions of 60%, 20%, and 20%, respectively. Here, we provide the detailed information of these datasets we used in our experiments.

AgeDB is a regression dataset on age prediction collected by (Moschoglou et al., 2017). It contains 16.4K face images with a minimum age of 0 and a maximum age of 101. We generated the interval regression dataset AgeDB-Interval at $q = 10, 20, 30, 40,$ and 50, respectively, and manually corrected the unreasonable intervals, such as intervals containing negative ages and intervals containing too old ages (less than 0 and greater than 150).

IMDB-WIKI is a regression dataset about age prediction collected by (Rothe et al., 2018). It contains 523.0K face images, and we filtered the images that do not match the criteria and finally kept 213.5K images, where the minimum age is 0 years and the maximum age is 186 years. We generated the interval regression dataset IMDB-WIKI-Interval at $q = 20, 30,$ and 40, respectively, and manually corrected the unreasonable intervals, such as those containing negative ages and those containing too old ages (less than 0 and greater than 200).

Semantic Textual Similarity Benchmark (STS-B) (Cer et al., 2017) is a collection of sentence pairs extracted from news headlines, video and image captions, and natural language inference data. Each sentence pair is scored for similarity by multiple annotators, and the final score is averaged as the final score. We created a dataset with 15.7K from (Yang et al., 2021). We generated interval regression datasets for STS-B-Interval at $q = 3.0, 3.5, 4.0, 4.5$ and 5.0, respectively.

We conducted experiments on 6 UCI benchmark datasets including Abalone, Airfoil, Auto-mpg, Housing, Concrete and Power-plant. All of these datasets can be downloaded from the UCI Machine Learning. Based on the span of the dataset labels, we selected two larger q values to generate interval regression data for each dataset.

E.2. Evaluation Metrics

We describe in detail all the evaluation metrics we used in our experiments.

MSE. The mean squared error (MSE) is defined as $\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$, where n denotes the number of samples, y_i denotes the ground truth value, and $\widehat{y}_i$ denotes the predicted value. MSE represents the averaged squared difference between the ground truth and predicted values over all samples.

MAE. The mean absolute error (MAE) is defined as $\frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i|$, where n denotes the number of samples, y_i denotes the ground truth value, and $\widehat{y}_i$ denotes the predicted value. MAE represents the averaged absolute difference between the ground truth and predicted values over all samples.

GM. We use the Geometric Mean (GM) proposed by (Yang et al., 2021) as our evaluation method, and is defined as $(\prod_{i=1}^{n} e_i)^{\frac{1}{n}}$, where $e_i \triangleq |y_i - \widehat{y}_i|$. GM is using the geometric mean to describe the fairness of the model predictions rather than the arithmetic mean.

Table 5. Complete evaluation results on AgeDB

Metric		MSE					MAE					GM				
Approach		$q=30$	$q=40$	$q=50$	$q=60$	$q=70$	$q=30$	$q=40$	$q=50$	$q=60$	$q=70$	$q=30$	$q=40$	$q=50$	$q=60$	$q=70$
Supervised				102.71 (3.12)					7.82 (0.14)					5.22 (0.13)		
LEFT	MAE	158.38 (6.71)	210.34 (19.85)	205.75 (22.04)	283.83 (15.36)	355.59 (105.16)	9.88 (0.22)	11.59 (0.61)	11.27 (0.80)	13.62 (0.62)	14.93 (2.38)	6.32 (0.13)	7.69 (0.32)	7.37 (0.67)	9.14 (0.40)	10.00 (1.77)
	MSE	134.99 (4.03)	196.83 (19.28)	221.54 (18.04)	295.49 (32.09)	347.27 (80.77)	9.25 (0.06)	11.19 (0.71)	11.91 (0.48)	13.90 (1.24)	14.98 (2.10)	6.08 (0.04)	7.46 (0.55)	7.88 (0.39)	9.54 (1.18)	10.47 (1.83)
	Huber	156.34 (17.95)	175.95 (7.61)	208.03 (18.36)	317.17 (22.57)	360.11 (53.12)	9.85 (0.57)	10.54 (0.19)	11.42 (0.61)	14.39 (0.37)	15.41 (1.30)	6.37 (0.33)	7.00 (0.14)	7.56 (0.59)	9.62 (0.05)	10.64 (1.16)
RIGHT	MAE	154.87 (13.12)	196.95 (20.96)	233.18 (26.58)	255.24 (15.12)	428.05 (41.79)	9.91 (0.45)	11.31 (0.58)	12.48 (0.79)	13.12 (0.40)	17.51 (0.93)	6.64 (0.39)	7.74 (0.40)	8.67 (0.65)	8.97 (0.29)	12.77 (0.94)
	MSE	146.85 (24.39)	215.06 (23.92)	260.37 (15.32)	304.71 (49.88)	452.61 (45.87)	9.57 (0.86)	11.96 (0.66)	13.27 (0.49)	14.47 (1.29)	18.15 (1.15)	6.34 (0.87)	8.04 (0.54)	9.12 (0.34)	10.28 (1.10)	13.31 (1.26)
	Huber	149.14 (7.74)	179.72 (10.57)	246.29 (16.02)	279.70 (17.45)	436.29 (78.90)	9.72 (0.31)	10.84 (0.37)	12.88 (0.48)	13.90 (0.44)	17.50 (1.86)	6.54 (0.19)	7.41 (0.25)	8.85 (0.63)	9.77 (0.24)	12.77 (1.79)
Middle	MAE	116.14 (2.57)	133.44 (5.05)	129.55 (1.37)	138.95 (5.22)	150.88 (3.66)	8.38 (0.13)	8.93 (0.15)	8.97 (0.13)	9.21 (0.11)	9.57 (0.12)	5.44 (0.04)	5.75 (0.04)	6.22 (0.72)	5.93 (0.08)	6.35 (0.11)
	MSE	119.90 (6.23)	133.27 (5.18)	128.84 (3.01)	138.36 (6.10)	149.82 (5.28)	8.45 (0.18)	8.94 (0.22)	8.89 (0.21)	9.32 (0.27)	9.53 (0.07)	5.57 (0.18)	5.91 (0.05)	5.75 (0.10)	6.16 (0.24)	6.33 (0.03)
	Huber	121.78 (4.75)	131.43 (3.84)	131.38 (2.20)	140.40 (6.18)	149.25 (0.70)	8.62 (0.14)	8.92 (0.15)	8.96 (0.08)	9.28 (0.19)	9.65 (0.09)	5.54 (0.07)	5.76 (0.10)	5.83 (0.05)	6.15 (0.18)	6.35 (0.12)
CRM		221.66 (1.45)	303.52 (11.12)	398.50 (14.80)	523.53 (4.63)	653.81 (17.30)	12.18 (0.07)	14.57 (0.30)	17.17 (0.40)	20.11 (0.08)	22.89 (0.09)	8.42 (0.07)	10.51 (0.24)	12.82 (0.44)	15.86 (0.20)	18.63 (0.21)
RANN		125.04 (1.09)	126.02 (1.74)	129.86 (1.00)	139.83 (3.41)	148.25 (2.33)	8.69 (0.04)	8.73 (0.07)	8.89 (0.05)	9.32 (0.22)	9.69 (0.11)	5.62 (0.12)	5.69 (0.07)	5.92 (0.03)	6.82 (1.01)	6.42 (0.09)
SINN		218.16 (1.53)	302.06 (6.53)	404.80 (2.70)	524.74 (7.17)	649.27 (2.94)	12.11 (0.08)	14.54 (0.24)	17.41 (0.04)	19.97 (0.13)	22.70 (0.14)	8.39 (0.14)	10.45 (0.29)	13.32 (0.13)	15.69 (0.24)	18.52 (0.25)
IN		118.85 (5.20)	133.86 (6.46)	138.58 (4.06)	147.95 (2.82)	152.34 (9.30)	8.46 (0.17)	9.00 (0.13)	9.15 (0.09)	9.53 (0.06)	9.65 (0.21)	5.75 (0.25)	5.88 (0.09)	5.94 (0.05)	6.23 (0.16)	6.30 (0.15)
LM		**115.76 (1.64)**	**121.24 (4.73)**	**123.51 (2.78)**	**128.04 (3.52)**	**129.47 (7.15)**	**8.36 (0.07)**	**8.67 (0.03)**	**8.75 (0.14)**	**8.82 (0.16)**	**8.92 (0.14)**	**5.41 (0.03)**	**5.61 (0.02)**	**5.65 (0.13)**	**5.73 (0.07)**	**5.84 (0.11)**

Table 6. All results on IMDB-WIKI.

Metrics		MSE					MAE					GM				
Approach		$q=40$	$q=50$	$q=60$	$q=70$	$q=80$	$q=40$	$q=50$	$q=60$	$q=70$	$q=80$	$q=40$	$q=50$	$q=60$	$q=70$	$q=80$
Supervised				123.82 (3.06)					8.38 (0.07)					5.40 (0.09)		
LEFT	MAE	270.15 (15.14)	336.73 (30.92)	402.05 (39.13)	494.84 (28.64)	581.54 (19.81)	13.06 (0.53)	14.76 (0.77)	16.28 (1.10)	18.30 (0.82)	20.39 (0.46)	8.60 (0.63)	10.09 (0.53)	13.54 (2.46)	12.98 (0.93)	15.28 (0.69)
	MSE	255.40 (13.45)	305.92 (7.33)	358.15 (25.61)	383.77 (16.82)	421.16 (94.72)	12.67 (0.40)	13.95 (0.60)	15.21 (0.57)	15.86 (0.36)	16.66 (2.36)	8.24 (0.43)	9.35 (0.29)	10.34 (0.42)	10.91 (0.45)	11.70 (2.36)
	Huber	291.25 (14.98)	320.37 (10.01)	385.99 (41.84)	512.73 (107.26)	596.39 (46.08)	13.39 (0.26)	14.24 (0.42)	15.82 (1.14)	18.57 (2.54)	20.71 (0.79)	8.71 (0.17)	11.40 (3.00)	10.83 (1.07)	13.46 (2.67)	15.57 (0.47)
RIGHT	MAE	197.26 (17.24)	265.60 (27.79)	328.77 (63.99)	608.51 (76.85)	679.26 (80.82)	11.15 (0.62)	13.38 (0.75)	14.74 (1.55)	20.80 (1.50)	22.82 (2.21)	8.07 (0.55)	10.74 (1.61)	13.29 (3.84)	15.69 (1.31)	17.55 (3.04)
	MSE	214.42 (10.40)	280.99 (12.52)	357.06 (64.59)	497.47 (43.98)	643.99 (85.80)	11.71 (0.35)	13.50 (0.35)	15.29 (1.54)	18.82 (1.05)	21.38 (1.49)	7.82 (0.31)	9.17 (0.32)	10.48 (1.18)	13.63 (0.96)	15.55 (1.24)
	Huber	198.97 (6.47)	243.81 (11.64)	379.34 (45.10)	536.87 (21.50)	491.39 (103.54)	11.23 (0.21)	12.82 (0.31)	16.19 (1.01)	19.63 (0.67)	18.24 (2.22)	8.48 (0.16)	10.84 (0.56)	14.15 (2.10)	14.14 (0.48)	12.83 (1.88)
Middle	MAE	140.29 (6.68)	148.93 (3.00)	152.20 (9.59)	154.79 (4.29)	157.17 (5.96)	8.96 (0.10)	9.24 (0.08)	9.47 (0.36)	9.55 (0.21)	9.76 (0.22)	5.94 (0.38)	6.20 (0.51)	6.82 (0.06)	7.43 (0.71)	6.96 (0.61)
	MSE	135.57 (2.77)	142.68 (2.43)	144.34 (2.60)	153.10 (9.34)	153.80 (3.57)	8.89 (0.10)	9.10 (0.06)	9.23 (0.11)	9.61 (0.30)	9.79 (0.08)	5.97 (0.54)	5.92 (0.20)	6.45 (0.70)	6.61 (0.16)	6.97 (0.43)
	Huber	142.21 (6.33)	148.52 (3.93)	153.97 (2.31)	152.52 (3.48)	154.77 (3.58)	9.04 (0.19)	9.22 (0.15)	9.49 (0.05)	9.49 (0.12)	9.70 (0.15)	5.71 (0.13)	6.05 (0.17)	6.25 (0.24)	6.12 (0.10)	6.41 (0.14)
CRM		302.50 (12.53)	380.12 (15.30)	519.7 (10.23)	602.65 (7.20)	740.99 (19.56)	14.38 (0.42)	16.81 (0.28)	20.11 (0.20)	21.85 (0.16)	24.61 (0.78)	10.18 (0.56)	12.81 (0.46)	13.26 (0.44)	18.16 (0.39)	21.34 (0.56)
RANN		137.37 (3.09)	140.79 (2.28)	145.40 (2.20)	150.11 (2.48)	166.33 (4.02)	8.98 (0.10)	8.98 (0.12)	9.32 (0.08)	9.52 (0.07)	10.09 (0.08)	5.82 (0.12)	6.14 (0.14)	6.74 (0.18)	6.85 (0.25)	6.63 (0.10)
SINN		314.49 (4.08)	385.29 (8.26)	515.80 (9.95)	629.51 (11.10)	754.28 (15.20)	14.79 (0.14)	16.73 (0.19)	19.84 (0.30)	22.36 (0.23)	24.73 (0.65)	10.88 (0.23)	12.59 (0.45)	15.77 (0.55)	17.31 (0.72)	20.28 (0.49)
IN		147.09 (0.59)	148.71 (1.01)	152.66 (2.54)	155.73 (5.67)	156.04 (2.89)	9.25 (0.04)	9.28 (0.09)	9.51 (0.10)	9.59 (0.17)	9.79 (0.06)	6.45 (0.24)	7.08 (0.36)	6.60 (0.42)	6.90 (0.70)	7.61 (0.15)
LM		**133.98 (1.57)**	**134.15 (2.49)**	**141.45 (2.32)**	**148.19 (2.44)**	**146.52 (5.00)**	**8.75 (0.06)**	**8.83 (0.07)**	**9.07 (0.11)**	**9.40 (0.08)**	**9.39 (0.04)**	**5.59 (0.08)**	**5.79 (0.20)**	**5.81 (0.07)**	**6.10 (0.10)**	**6.12 (0.14)**

Pearson correlation. Pearson correlation is an evaluation of the linear relationship between the predicted value and the ground truth value, and is defined as $\frac{\sum_{i=1}^{n}(y_i-\overline{y})(\widehat{y}_i-\overline{\widehat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}\sqrt{\sum_{i=1}^{n}(\widehat{y}_i-\overline{\widehat{y}})^2}}$, where $\overline{y}$ denotes the average of all ground truth values, $\overline{\widehat{y}}$ denotes the average of all predicted values, i.e., $\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$, $\overline{\widehat{y}}=\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_i$.

F. Additional Results

In the experiment of main paper, we show some of the experiments on the three datasets AgeDB, IMDB-WIKI and STS-B. Here, we provide the complete evaluation results on the nine used datasets, which include more evaluation metrics in

Table 7. All results on STS-B.

Metrics		MSE					MAE					Pearson				
Approach		$q=3.0$	$q=3.5$	$q=4.0$	$q=4.5$	$q=5.0$	$q=3.0$	$q=3.5$	$q=4.0$	$q=4.5$	$q=5.0$	$q=3.0$	$q=3.5$	$q=4.0$	$q=4.5$	$q=5.0$
Supervised				1.16 (0.05)					0.87 (0.02)					0.71 (0.01)		
LEFT	MAE	1.54 (0.06)	1.88 (0.06)	2.35 (0.15)	2.63 (0.13)	3.03 (0.06)	1.01 (0.02)	1.11 (0.02)	1.26 (0.05)	1.33 (0.03)	1.43 (0.02)	**0.69** **(0.01)**	**0.66** **(0.01)**	0.64 (0.01)	0.62 (0.02)	0.59 (0.01)
	MSE	2.02 (0.08)	2.16 (0.07)	2.55 (0.12)	2.88 (0.11)	3.13 (0.12)	1.17 (0.03)	1.20 (0.02)	1.32 (0.03)	1.40 (0.03)	1.47 (0.04)	0.66 (0.01)	0.65 (0.01)	0.63 (0.01)	0.61 (0.01)	0.58 (0.03)
	Huber	1.92 (0.13)	2.26 (0.10)	2.51 (0.14)	2.76 (0.10)	3.05 (0.15)	1.14 (0.04)	1.23 (0.03)	1.31 (0.05)	1.37 (0.03)	1.44 (0.04)	0.67 (0.01)	0.65 (0.01)	0.63 (0.01)	0.61 (0.01)	0.59 (0.02)
RIGHT	MAE	1.75 (0.14)	1.87 (0.19)	2.09 (0.18)	2.05 (0.28)	2.31 (0.40)	1.06 (0.04)	1.12 (0.07)	1.17 (0.05)	1.16 (0.08)	1.23 (0.10)	0.68 (0.01)	0.63 (0.05)	0.62 (0.02)	0.55 (0.09)	0.55 (0.05)
	MSE	1.53 (0.05)	1.74 (0.13)	1.78 (0.19)	2.00 (0.17)	2.26 (0.16)	1.00 (0.02)	1.06 (0.03)	1.08 (0.06)	1.14 (0.05)	1.21 (0.03)	0.63 (0.01)	0.60 (0.02)	0.56 (0.05)	0.53 (0.02)	0.52 (0.02)
	Huber	1.56 (0.08)	1.77 (0.11)	1.86 (0.27)	1.91 (0.15)	2.39 (0.08)	1.01 (0.03)	1.07 (0.04)	1.10 (0.08)	1.12 (0.05)	1.24 (0.03)	0.64 (0.02)	0.62 (0.01)	0.58 (0.02)	0.56 (0.02)	0.52 (0.02)
Middle	MAE	**1.18** **(0.03)**	1.23 (0.05)	**1.27** **(0.07)**	**1.31** **(0.03)**	1.41 (0.05)	**0.88** **(0.01)**	0.91 (0.02)	**0.92** **(0.03)**	**0.94** **(0.01)**	0.98 (0.02)	**0.69** **(0.01)**	**0.66** **(0.02)**	**0.66** **(0.01)**	**0.64** **(0.01)**	**0.62** **(0.02)**
	MSE	1.31 (0.02)	1.33 (0.05)	1.41 (0.04)	1.41 (0.01)	1.49 (0.03)	0.94 (0.01)	0.95 (0.02)	0.98 (0.01)	0.99 (0.01)	1.02 (0.01)	0.65 (0.01)	0.64 (0.02)	0.62 (0.01)	0.61 (0.01)	0.59 (0.01)
	Huber	1.37 (0.16)	1.30 (0.04)	1.37 (0.08)	1.39 (0.02)	1.48 (0.03)	0.96 (0.06)	0.94 (0.02)	0.97 (0.03)	0.97 (0.01)	1.01 (0.01)	0.64 (0.03)	0.64 (0.01)	0.62 (0.01)	0.62 (0.01)	0.59 (0.01)
CRM		1.36 (0.02)	1.50 (0.10)	1.53 (0.08)	1.52 (0.06)	1.56 (0.06)	0.96 (0.01)	1.00 (0.04)	1.02 (0.02)	1.02 (0.02)	1.05 (0.03)	0.64 (0.01)	0.60 (0.03)	0.59 (0.03)	0.58 (0.03)	0.56 (0.02)
RANN		1.44 (0.05)	1.52 (0.04)	1.55 (0.06)	1.63 (0.06)	1.62 (0.06)	0.98 (0.02)	1.01 (0.02)	1.02 (0.03)	1.05 (0.02)	1.06 (0.03)	0.59 (0.01)	0.57 (0.02)	0.55 (0.02)	0.52 (0.02)	0.51 (0.02)
SINN		1.37 (0.04)	1.50 (0.05)	1.50 (0.10)	1.54 (0.05)	1.61 (0.09)	0.96 (0.01)	1.00 (0.01)	1.01 (0.03)	1.03 (0.02)	1.06 (0.03)	0.64 (0.01)	0.60 (0.01)	0.60 (0.02)	0.58 (0.01)	0.54 (0.03)
IN		1.26 (0.12)	1.39 (0.21)	1.37 (0.14)	1.37 (0.08)	1.41 (0.07)	0.91 (0.04)	0.96 (0.06)	0.96 (0.04)	0.96 (0.03)	0.98 (0.03)	0.67 (0.04)	0.62 (0.06)	0.62 (0.05)	0.62 (0.01)	0.61 (0.01)
LM		**1.19** **(0.04)**	**1.22** **(0.02)**	**1.27** **(0.04)**	**1.31** **(0.04)**	**1.37** **(0.05)**	**0.88** **(0.01)**	**0.90** **(0.01)**	0.93 (0.01)	**0.94** **(0.01)**	**0.96** **(0.01)**	**0.69** **(0.01)**	**0.66** **(0.01)**	0.65 (0.01)	**0.64** **(0.02)**	**0.62** **(0.02)**

addition to the results in the main paper.

F.1. Complete Results on AgeDB

We show the complete results of AgeDB in Table 5. In the table, we show the test performance (mean and std) of each method with ResNet-50, evaluated using MSE, MAE and GM. We repeat the sampling-and-training process 3 times. The best performance is highlighted in bold. In addition, Our Vs Sup means the mean error between our proposed method and the supervised method (Our - Supervised). Specifically, we use red color to indicate that our method is weaker than the supervised method and green color to indicate that our method is better than the supervised method. As the table shows, our proposed method LM has significant advantages in all evaluation metrics.

F.2. Complete Results on IMDB-WIKI

We show the complete results of IMDB-WIKI in Table 6. Similar to AgeDB, we evaluate each method using MSE, MAE and GM. We repeat the sampling-and-training process 3 times. It is worth noting that we chose different q values for AgeDB and IMDB-WIKI, although both are age predictions. IMDB-WIKI has a larger training set, and we want to test our method on a large dataset in a harsh environment (large q). As shown in the table 6, our proposed method LM has significant advantages in all evaluation metrics. Even in the case of $q = 80$, the performance does not degrade excessively after learning from a large number of training sets.

F.3. Complete Results on STS-B

We show the complete results of STS-B in Table 7. In the table, we show the test performance (mean and std) of each method with BiLSTM + GloVe word embeddings baseline, evaluated using MSE, MAE and Pearson. Unlike AgeDB and IMDB-WIKI, STS-B has a smaller span of labels, so we can only choose smaller q values to generate interval data. As shown in the table, the difference between methods is not significant when q is small. As q keeps increasing, all the methods tend to decrease in performance, while our method decreases more slowly and has a significant advantage at large q.

F.4. Complete Results on UCI Benchmark Datasets

Table 10 and Table 11 show the mean squared error with standard deviation on the test set using the MLP model and the linear model, respectively. We evaluate each method using MSE, MAE, Pearson correlation and GM. We repeat the sampling-and-training process 5 times. As the table shows, our proposed method LM has significant advantages in all

Table 8. Comparison of MMIT and LM with MLP model

Metrics		MSE			MAE			Pearson			GM		
Approach		Linear	Square	LM_MLP	Linear	Square	LM_MLP	Linear	Square	LM_MLP	Linear	Square	LM_MLP
Abalone	q=30	6.41 (0.31)	6.77 (0.82)	**4.66 (0.49)**	1.88 (0.03)	1.98 (0.15)	**1.50 (0.05)**	0.63 (0.02)	0.63 (0.04)	**0.74 (0.02)**	1.17 (0.03)	1.28 (0.09)	**0.91 (0.07)**
	q=40	7.03 (0.79)	6.99 (0.69)	**4.81 (0.42)**	2.01 (0.17)	2.02 (0.12)	**1.53 (0.08)**	0.61 (0.02)	0.62 (0.01)	**0.75 (0.01)**	1.36 (0.13)	1.37 (0.13)	**0.90 (0.04)**
Airfoil	q=30	18.74 (3.41)	17.98 (1.70)	**16.72 (3.42)**	3.33 (0.30)	3.25 (0.14)	**3.09 (0.35)**	0.80 (0.05)	0.80 (0.04)	**0.81 (0.04)**	2.15 (0.22)	2.03 (0.12)	**1.93 (0.27)**
	q=40	23.61 (3.20)	19.57 (1.98)	**18.31 (2.63)**	3.72 (0.26)	3.46 (0.16)	**3.24 (0.27)**	0.74 (0.03)	0.78 (0.02)	**0.80 (0.03)**	2.38 (0.25)	2.25 (0.16)	**2.08 (0.22)**
Auto-mpg	q=30	17.82 (2.72)	15.51 (2.23)	**9.63 (1.62)**	3.16 (0.25)	2.91 (0.25)	**2.19 (0.18)**	0.84 (0.04)	0.87 (0.01)	**0.92 (0.01)**	1.95 (0.23)	1.81 (0.25)	**1.36 (0.16)**
	q=40	19.07 (4.45)	25.29 (5.37)	**11.11 (3.03)**	3.24 (0.36)	3.78 (0.44)	**2.31 (0.30)**	0.85 (0.03)	0.80 (0.05)	**0.92 (0.01)**	2.00 (0.25)	2.50 (0.38)	**1.40 (0.12)**
Housing	q=30	32.85 (8.39)	32.00 (9.79)	**22.13 (3.71)**	3.94 (0.32)	3.78 (0.38)	**3.47 (0.57)**	0.80 (0.05)	0.82 (0.05)	**0.86 (0.04)**	2.36 (0.12)	2.31 (0.13)	**2.08 (0.32)**
	q=40	30.67 (5.06)	31.79 (5.00)	**24.53 (8.00)**	3.99 (0.49)	4.11 (0.38)	**3.49 (0.48)**	0.82 (0.04)	0.81 (0.04)	**0.85 (0.05)**	2.56 (0.51)	2.61 (0.32)	**2.22 (0.31)**
Concrete	q=70	106.74 (8.94)	105.08 (14.70)	**58.45 (3.52)**	8.02 (0.33)	8.00 (0.53)	**5.86 (0.41)**	0.80 (0.03)	0.80 (0.02)	**0.89 (0.01)**	5.21 (0.26)	5.16 (0.50)	**3.68 (0.32)**
	q=80	107.80 (13.61)	109.25 (11.09)	**59.47 (7.81)**	8.01 (0.41)	8.03 (0.51)	**5.71 (0.34)**	0.80 (0.04)	0.80 (0.02)	**0.89 (0.02)**	5.25 (0.25)	5.24 (0.47)	**3.57 (0.27)**

Table 9. Comparison of AVGL_MSE and AVGL_MAE

Metric		MSE		MAE		Pearson		GM	
Approach		AVGL_MSE	AVGL_MAE	AVGL_MSE	AVGL_MAE	AVGL_MSE	AVGL_MAE	AVGL_MSE	AVGL_MAE
Abalone	$q=30$	6.41 (0.43)	**4.66 (0.49)**	1.94 (0.06)	**1.50 (0.05)**	0.73 (0.03)	**0.74 (0.02)**	1.04 (0.05)	**0.91 (0.07)**
	$q=40$	7.77 (0.68)	**4.81 (0.42)**	1.96 (0.08)	**1.53 (0.08)**	0.73 (0.01)	**0.75 (0.01)**	1.18 (0.05)	**0.90 (0.04)**
Airfoil	$q=30$	19.24 (1.70)	**16.72 (3.42)**	3.41 (0.24)	**3.09 (0.35)**	0.78 (0.04)	**0.81 (0.04)**	2.21 (0.24)	**1.93 (0.27)**
	$q=40$	24.37 (1.12)	**18.31 (2.63)**	3.99 (0.11)	**3.24 (0.27)**	0.77 (0.01)	**0.80 (0.03)**	2.74 (0.21)	**2.08 (0.22)**
Auto-mpg	$q=30$	11.63 (1.49)	**9.63 (1.62)**	2.50 (0.29)	**2.19 (0.18)**	**0.92 (0.01)**	0.92 (0.01)	1.75 (0.17)	**1.36 (0.16)**
	$q=40$	20.08 (5.51)	**11.11 (3.03)**	3.43 (0.49)	**2.31 (0.30)**	0.91 (0.02)	**0.92 (0.01)**	2.42 (0.43)	**1.40 (0.12)**
Housing	$q=30$	25.69 (10.69)	**22.13 (3.71)**	**3.44 (0.43)**	3.47 (0.57)	0.85 (0.05)	**0.86 (0.04)**	2.27 (0.16)	**2.08 (0.32)**
	$q=40$	31.74 (4.97)	**24.53 (8.00)**	4.07 (0.39)	**3.49 (0.48)**	0.83 (0.03)	**0.85 (0.05)**	2.55 (0.43)	**2.22 (0.31)**
Concrete	$q=70$	69.04 (7.92)	**58.45 (3.52)**	6.55 (0.34)	**5.86 (0.41)**	0.88 (0.02)	**0.89 (0.01)**	4.26 (0.38)	**3.68 (0.32)**
	$q=80$	86.05 (7.66)	**59.47 (7.81)**	7.29 (0.26)	**5.71 (0.34)**	0.86 (0.03)	**0.89 (0.02)**	5.02 (0.14)	**3.57 (0.27)**
Power-plant	$q=60$	31.70 (2.89)	**23.54 (1.41)**	4.52 (0.19)	**3.84 (0.11)**	0.95 (0.00)	**0.96 (0.00)**	3.26 (0.19)	**2.62 (0.03)**
	$q=70$	47.59 (2.86)	**24.91 (0.60)**	5.55 (0.13)	**3.96 (0.05)**	0.95 (0.00)	**0.96 (0.00)**	4.07 (0.29)	**2.76 (0.14)**

evaluation metrics. In particular, by comparing the experimental results reported in Table 10 and Table 11 we can observe that training with the MLP model is generally better than training with the linear model. This observation is consistent with the common knowledge that MLP models are more powerful than linear models. It is worth noting that we also compare with the maximum margin interval trees method (MMIT(Drouin et al., 2017)), which is similar to our limiting method. They both want to limit the predicted values to the interval. Table 8 shows the results of our method with the MLP model and MMIT with linear (Linear) and squared (Square) hinge loss variants.

F.5. Comparison between AVGL_MSE and AVGL_MAE

Table 9 shows the comparison between AVGL_MSE and AVGL_MAE on the used UCI benchmark datasets with the MLP model, where AVGL_MSE and AVGL_MAE substitute the regression loss function in the average loss method with MSE and MAE, respectively. As shown in the table, AVGL_MAE is significantly better than AVGL_MSE.

Table 10. Complete evaluation results on UCI benchmark datasets with MLP model

Datasets	Metrese	q	Supervised	Leftmost MSE	Leftmost MAE	Leftmost Huber	Rightmost MSE	Rightmost MAE	Rightmost Huber	Middlemost MSE	Middlemost MAE	Middlemost Huber	CRM	RANN	SINN	IN	LM
Abalone	MSE	30	4.57 (0.50)	65.33 (2.09)	55.73 (1.79)	59.82 (3.07)	7.99 (0.70)	8.08 (0.71)	8.03 (0.71)	6.38 (0.44)	5.45 (0.43)	6.01 (0.44)	6.51 (0.52)	6.67 (0.48)	6.44 (0.44)	5.61 (0.40)	**4.66 (0.49)**
		40		90.12 (3.16)	85.71 (6.28)	89.27 (4.47)	8.13 (0.67)	8.23 (0.63)	8.21 (0.64)	7.77 (0.68)	7.81 (0.76)	7.79 (0.76)	7.85 (0.81)	7.74 (0.64)	7.85 (0.81)	7.84 (0.75)	**4.81 (0.42)**
	MAE	30	1.46 (0.04)	7.60 (0.11)	6.96 (0.13)	7.21 (0.16)	2.01 (0.08)	2.11 (0.08)	2.10 (0.08)	1.94 (0.06)	1.77 (0.08)	1.91 (0.07)	1.94 (0.07)	1.91 (0.07)	1.94 (0.07)	1.86 (0.05)	**1.50 (0.05)**
		40		9.03 (0.15)	8.80 (0.32)	9.03 (0.18)	2.02 (0.08)	2.12 (0.07)	2.11 (0.07)	1.96 (0.08)	1.95 (0.06)	1.96 (0.06)	1.99 (0.07)	2.02 (0.08)	1.99 (0.07)	1.97 (0.06)	**1.53 (0.08)**
	Pearson	30	0.75 (0.02)	0.73 (0.02)	0.72 (0.02)	0.73 (0.02)	0.71 (0.02)	0.71 (0.02)	0.71 (0.02)	0.73 (0.02)	0.72 (0.02)	0.72 (0.03)	0.73 (0.02)	0.72 (0.03)	0.73 (0.02)	0.72 (0.02)	**0.74 (0.02)**
		40		0.72 (0.02)	0.66 (0.03)	0.69 (0.03)	0.72 (0.01)	0.71 (0.03)	0.71 (0.02)	0.73 (0.01)	0.73 (0.01)	0.73 (0.01)	0.73 (0.00)	0.73 (0.00)	0.73 (0.00)	0.73 (0.01)	**0.75 (0.01)**
	GM	30	0.85 (0.03)	7.18 (0.11)	6.51 (0.12)	6.79 (0.16)	1.18 (0.05)	1.24 (0.05)	1.25 (0.05)	1.04 (0.05)	1.06 (0.08)	1.03 (0.05)	1.14 (0.15)	1.19 (0.18)	1.05 (0.05)	1.04 (0.03)	**0.91 (0.07)**
		40		8.62 (0.17)	8.31 (0.27)	8.56 (0.15)	1.18 (0.06)	1.28 (0.04)	1.28 (0.04)	1.18 (0.05)	1.14 (0.04)	1.18 (0.05)	1.20 (0.09)	1.20 (0.07)	1.20 (0.09)	1.18 (0.06)	**0.90 (0.04)**
Airfoil	MSE	30	16.27 (3.23)	122.58 (89.68)	89.68 (5.26)	93.90 (8.01)	105.20 (9.19)	79.28 (7.75)	85.16 (14.44)	19.24 (1.70)	18.71 (1.16)	18.24 (2.02)	17.40 (3.07)	17.30 (2.86)	17.35 (3.13)	19.20 (1.30)	**16.72 (3.42)**
		40		200.70 (13.62)	164.64 (9.28)	183.30 (23.42)	134.36 (13.66)	115.92 (5.56)	120.34 (7.57)	24.32 (1.16)	20.43 (1.87)	20.99 (1.89)	23.12 (2.30)	22.98 (1.92)	23.08 (2.38)	19.37 (2.51)	**18.31 (2.63)**
	MAE	30	3.02 (0.33)	9.81 (0.52)	8.50 (0.39)	8.71 (0.62)	8.00 (0.26)	7.78 (0.38)	7.98 (0.51)	3.41 (0.24)	3.30 (0.15)	3.28 (0.25)	3.26 (0.33)	3.23 (0.32)	3.26 (0.32)	3.33 (0.16)	**3.09 (0.35)**
		40		11.93 (0.31)	11.96 (0.38)	12.18 (0.60)	8.93 (0.64)	9.24 (0.43)	9.29 (0.46)	3.97 (0.12)	3.54 (0.13)	3.61 (0.20)	3.85 (0.22)	3.85 (0.22)	3.85 (0.21)	3.47 (0.26)	**3.24 (0.27)**
	Pearson	30	0.81 (0.05)	0.78 (0.04)	0.78 (0.04)	0.78 (0.04)	0.78 (0.04)	0.77 (0.03)	0.79 (0.05)	0.79 (0.04)	0.78 (0.03)	0.79 (0.03)	**0.81 (0.05)**	**0.81 (0.05)**	0.80 (0.05)	0.78 (0.04)	**0.81 (0.04)**
		40		0.77 (0.02)	0.75 (0.03)	0.76 (0.03)	0.76 (0.02)	0.76 (0.02)	0.77 (0.02)	0.78 (0.01)	0.78 (0.03)	0.79 (0.02)	0.79 (0.02)	0.79 (0.02)	0.79 (0.02)	**0.80 (0.04)**	0.80 (0.03)
	GM	30	1.90 (0.22)	6.59 (0.35)	7.29 (0.60)	7.23 (0.48)	5.42 (0.73)	5.42 (0.37)	5.43 (0.33)	2.17 (0.27)	2.12 (0.19)	2.06 (0.24)	2.11 (0.21)	2.12 (0.25)	2.11 (0.20)	2.21 (0.28)	**1.93 (0.27)**
		40		8.98 (0.30)	8.66 (0.84)	8.59 (0.91)	5.60 (0.35)	6.31 (1.01)	6.44 (0.89)	2.73 (0.21)	2.46 (0.07)	2.47 (0.22)	2.66 (0.09)	2.75 (0.25)	2.64 (0.12)	2.26 (0.19)	**2.08 (0.22)**
Auto-mpg	MSE	30	9.73 (1.95)	76.60 (12.72)	34.99 (12.83)	50.43 (18.98)	51.52 (20.31)	30.41 (11.25)	24.06 (7.60)	11.59 (1.42)	11.69 (1.69)	11.52 (1.17)	11.68 (1.25)	23.60 (3.86)	11.67 (1.28)	11.60 (1.23)	**9.63 (1.95)**
		40		145.11 (19.52)	93.80 (23.82)	92.64 (32.76)	119.54 (20.53)	48.03 (15.25)	61.49 (28.88)	20.14 (5.44)	18.30 (4.52)	19.25 (4.69)	21.11 (6.18)	34.57 (8.40)	21.08 (6.30)	18.23 (4.97)	**11.11 (3.03)**
	MAE	30	2.16 (0.22)	7.97 (0.76)	4.47 (1.03)	5.98 (1.34)	6.33 (1.63)	4.47 (0.94)	3.89 (0.78)	2.52 (0.23)	2.51 (0.18)	2.48 (0.18)	2.50 (0.19)	3.76 (0.38)	2.50 (0.19)	2.50 (0.17)	**2.19 (0.18)**
		40		11.59 (1.02)	8.09 (1.16)	8.16 (1.84)	10.91 (1.75)	5.78 (1.03)	6.68 (1.92)	3.46 (0.48)	3.15 (0.37)	3.18 (0.43)	3.39 (0.50)	4.62 (0.59)	3.36 (0.52)	3.24 (0.41)	**2.31 (0.30)**
	Pearson	30	0.92 (0.02)	0.91 (0.02)	0.90 (0.02)	0.90 (0.02)	0.91 (0.01)	0.90 (0.01)	0.91 (0.01)	**0.92 (0.01)**	0.91 (0.02)	0.91 (0.01)	**0.92 (0.01)**	0.79 (0.05)	**0.92 (0.01)**	0.90 (0.01)	**0.92 (0.01)**
		40		0.91 (0.02)	0.90 (0.03)	0.91 (0.03)	0.89 (0.01)	0.87 (0.04)	0.87 (0.04)	**0.92 (0.02)**	0.91 (0.02)	**0.92 (0.02)**	0.91 (0.02)	0.72 (0.07)	0.91 (0.02)	0.91 (0.02)	**0.92 (0.01)**
	GM	30	1.31 (0.13)	6.72 (0.27)	2.82 (0.83)	5.05 (2.15)	4.69 (1.27)	3.17 (0.89)	2.60 (0.75)	1.75 (0.16)	1.60 (0.18)	1.57 (0.14)	1.60 (0.14)	2.55 (0.54)	1.57 (0.18)	1.60 (0.17)	**1.36 (0.16)**
		40		8.45 (1.50)	6.68 (1.87)	7.19 (2.52)	8.87 (1.08)	4.38 (0.99)	4.92 (1.45)	2.39 (0.51)	1.90 (0.20)	1.92 (0.20)	2.37 (0.59)	3.31 (0.40)	2.11 (0.42)	2.10 (0.34)	**1.40 (0.12)**
Housing	MSE	30	16.76 (6.28)	55.46 (8.52)	52.66 (15.47)	52.79 (7.06)	75.37 (27.20)	78.40 (8.87)	83.83 (15.28)	27.49 (11.60)	26.24 (8.20)	25.07 (5.82)	27.58 (6.80)	25.45 (8.95)	26.32 (10.12)	25.06 (7.59)	**22.13 (3.71)**
		40		88.85 (22.73)	101.70 (13.36)	83.96 (15.88)	109.01 (32.28)	124.34 (12.16)	124.14 (12.18)	30.70 (4.04)	35.84 (2.72)	32.63 (4.86)	32.66 (4.47)	31.15 (5.46)	34.92 (4.47)	33.00 (5.71)	**24.53 (8.00)**
	MAE	30	2.77 (0.44)	6.02 (0.65)	5.61 (1.06)	5.49 (0.61)	6.88 (0.54)	7.63 (0.33)	8.31 (0.85)	3.52 (0.36)	3.54 (0.45)	3.50 (0.27)	3.52 (0.50)	3.51 (0.56)	3.63 (0.66)	3.56 (0.43)	**3.47 (0.57)**
		40		7.15 (0.66)	7.46 (0.53)	7.17 (0.62)	7.87 (0.92)	8.27 (0.42)	8.26 (0.42)	4.06 (0.45)	4.41 (0.28)	4.11 (0.31)	4.14 (0.45)	3.95 (0.48)	4.25 (0.39)	4.19 (0.46)	**3.49 (0.48)**
	Pearson	30	0.91 (0.00)	0.84 (0.04)	0.83 (0.03)	0.83 (0.04)	0.84 (0.06)	0.83 (0.05)	0.81 (0.04)	0.85 (0.04)	0.85 (0.04)	0.85 (0.03)	0.84 (0.05)	0.85 (0.06)	0.85 (0.06)	0.85 (0.05)	**0.86 (0.04)**
		40		0.84 (0.04)	0.76 (0.02)	0.80 (0.03)	0.77 (0.05)	0.72 (0.04)	0.74 (0.05)	0.82 (0.03)	0.78 (0.03)	0.82 (0.03)	0.81 (0.03)	0.83 (0.03)	0.81 (0.03)	0.81 (0.06)	**0.85 (0.05)**
	GM	30	1.73 (0.21)	4.00 (0.50)	3.51 (0.90)	3.46 (0.62)	4.85 (0.54)	5.04 (0.41)	5.03 (0.44)	2.72 (0.11)	2.19 (0.24)	2.22 (0.15)	2.20 (0.28)	2.25 (0.31)	2.12 (0.24)	2.16 (0.21)	**2.08 (0.32)**
		40		4.92 (0.89)	4.78 (0.37)	4.69 (0.68)	5.04 (0.66)	5.06 (0.60)	4.93 (0.69)	2.41 (0.29)	2.67 (0.30)	2.41 (0.27)	2.74 (0.54)	2.42 (0.49)	2.49 (0.29)	2.60 (0.60)	**2.22 (0.31)**
Concrete	MSE	70	43.53 (1.00)	283.72 (13.93)	291.78 (17.85)	296.00 (7.37)	278.35 (18.24)	278.30 (18.16)	278.26 (18.18)	70.94 (8.19)	72.84 (7.78)	73.01 (6.65)	74.68 (5.80)	74.16 (4.23)	71.68 (5.84)	67.22 (5.82)	**58.45 (3.52)**
		80		420.44 (68.96)	424.03 (57.64)	423.70 (40.60)	284.03 (15.41)	284.04 (15.41)	284.02 (15.43)	88.26 (3.13)	95.09 (8.56)	86.54 (13.11)	86.40 (3.13)	80.77 (2.78)	83.38 (3.13)	92.48 (16.38)	**59.47 (7.81)**
	MAE	70	4.97 (0.14)	13.93 (1.04)	14.07 (0.99)	14.42 (0.64)	13.38 (0.49)	13.38 (0.49)	13.38 (0.49)	6.70 (0.54)	6.81 (0.25)	6.83 (0.32)	6.78 (0.81)	6.51 (0.82)	6.77 (0.82)	6.38 (0.34)	**5.86 (0.41)**
		80		17.74 (1.55)	17.67 (1.48)	17.93 (1.26)	13.47 (0.38)	13.47 (0.38)	13.47 (0.38)	7.42 (0.46)	7.62 (0.89)	7.35 (0.61)	7.40 (0.68)	7.26 (0.33)	7.48 (0.47)	7.49 (0.70)	**5.71 (0.34)**
	Pearson	70	0.92 (0.00)	0.86 (0.04)	0.84 (0.04)	0.86 (0.02)	0.86 (0.03)	0.82 (0.05)	0.85 (0.03)	0.88 (0.03)	0.88 (0.01)	0.87 (0.01)	0.87 (0.01)	0.88 (0.01)	0.87 (0.01)	0.88 (0.01)	**0.89 (0.01)**
		80		0.84 (0.03)	0.84 (0.03)	0.84 (0.04)	0.85 (0.02)	0.78 (0.04)	0.79 (0.06)	0.86 (0.03)	0.84 (0.03)	0.85 (0.03)	0.85 (0.03)	0.48 (0.03)	0.85 (0.03)	0.84 (0.03)	**0.89 (0.02)**
	GM	70	3.16 (0.27)	9.77 (1.24)	10.37 (1.41)	10.17 (0.94)	8.92 (0.47)	8.95 (0.49)	8.95 (0.48)	4.44 (0.38)	4.38 (0.36)	4.50 (0.30)	4.56 (0.30)	4.24 (0.34)	4.42 (0.36)	3.97 (0.14)	**3.68 (0.32)**
		80		13.14 (1.51)	13.14 (1.10)	13.49 (0.66)	9.33 (0.54)	9.34 (0.56)	9.34 (0.55)	5.16 (0.62)	5.09 (0.36)	4.90 (0.54)	4.96 (0.37)	4.91 (0.27)	5.01 (0.38)	5.08 (0.82)	**3.57 (0.27)**
Power-plant	MSE	60	22.23 (0.96)	264.98 (49.48)	201.00 (51.72)	245.19 (35.67)	262.00 (107.86)	246.15 (69.44)	268.46 (53.87)	31.96 (2.69)	28.95 (3.03)	29.38 (3.76)	32.85 (3.78)	31.30 (2.84)	32.66 (3.95)	29.36 (3.41)	**23.54 (1.41)**
		70		279.48 (3.14)	316.33 (72.10)	337.53 (61.63)	280.29 (76.53)	269.03 (81.76)	293.51 (2.20)	48.82 (1.53)	39.41 (2.76)	40.83 (3.41)	48.53 (1.20)	47.62 (2.37)	48.62 (1.22)	40.04 (3.13)	**24.91 (0.60)**
	MAE	60	3.71 (0.06)	14.48 (0.10)	12.55 (1.58)	14.08 (0.34)	13.92 (3.31)	13.98 (2.39)	14.60 (1.43)	4.55 (0.18)	4.30 (0.21)	4.36 (0.32)	4.59 (0.28)	4.49 (0.21)	4.58 (0.31)	4.37 (0.29)	**3.84 (0.11)**
		70		14.29 (0.16)	14.48 (1.51)	14.70 (0.94)	14.17 (1.85)	14.70 (2.56)	14.99 (0.10)	5.59 (0.14)	4.94 (0.18)	5.13 (0.20)	5.62 (0.08)	5.51 (0.10)	5.62 (0.07)	5.01 (0.16)	**3.96 (0.05)**
	Pearson	60	0.96 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	**0.96 (0.00)**	**0.96 (0.00)**	**0.96 (0.00)**	**0.96 (0.00)**	**0.96 (0.00)**	**0.96 (0.00)**	**0.96 (0.00)**
		70		0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	**0.96 (0.00)**
	GM	60	2.57 (0.14)	10.43 (1.26)	9.24 (0.89)	9.75 (0.18)	11.81 (2.43)	10.91 (1.87)	11.63 (0.91)	3.19 (0.14)	3.00 (0.23)	2.98 (0.17)	3.17 (0.24)	3.19 (0.14)	3.10 (0.20)	3.05 (0.11)	**2.62 (0.03)**
		70		9.84 (0.30)	9.45 (0.55)	9.65 (0.29)	10.82 (2.78)	11.35 (1.92)	12.80 (2.52)	3.95 (0.13)	3.56 (0.07)	3.56 (0.06)	3.85 (0.08)	3.89 (0.18)	3.87 (0.08)	3.52 (0.05)	**2.76 (0.14)**

Table 11. Complete evaluation results on UCI benchmark datasets with Linear model

Datasets	Metrese	q	Supervised	Leftmost MSE	Leftmost MAE	Leftmost Huber	Rightmost MSE	Rightmost MAE	Rightmost Huber	Middlemost MSE	Middlemost MAE	Middlemost Huber	CRM	RANN	SINN	IN	LM
Abalone	MSE	30	4.87 (0.51)	67.82 (3.53)	60.42 (4.52)	64.36 (3.58)	8.11 (0.68)	8.26 (0.59)	8.29 (0.59)	6.63 (0.44)	5.65 (0.48)	6.13 (0.44)	6.65 (0.45)	6.70 (0.45)	6.65 (0.45)	5.70 (0.49)	**4.92 (0.50)**
		40		93.18 (6.04)	92.28 (5.87)	93.58 (6.28)	8.33 (0.64)	8.33 (0.56)	8.37 (0.56)	7.80 (0.80)	7.80 (0.76)	7.77 (0.72)	8.02 (0.84)	7.54 (0.65)	8.02 (0.84)	8.44 (0.92)	**5.05 (0.50)**
	MAE	30	1.57 (0.06)	7.75 (0.20)	7.25 (0.28)	7.52 (0.15)	2.09 (0.08)	2.17 (0.07)	2.18 (0.07)	1.90 (0.06)	1.83 (0.08)	1.91 (0.06)	1.91 (0.06)	1.90 (0.08)	1.91 (0.06)	1.86 (0.08)	**1.58 (0.06)**
		40		9.17 (0.28)	9.17 (0.33)	9.22 (0.33)	2.11 (0.07)	2.15 (0.05)	2.16 (0.05)	1.95 (0.08)	1.93 (0.08)	1.94 (0.08)	1.93 (0.10)	1.94 (0.09)	1.93 (0.10)	1.94 (0.12)	**1.59 (0.09)**
	Pearson	30	0.73 (0.02)	0.72 (0.03)	0.71 (0.03)	0.71 (0.03)	0.71 (0.02)	0.70 (0.02)	0.70 (0.02)	0.72 (0.02)	0.71 (0.03)	0.71 (0.03)	0.72 (0.03)	0.72 (0.02)	0.72 (0.03)	0.72 (0.02)	**0.73 (0.02)**
		40		0.72 (0.02)	0.71 (0.03)	0.72 (0.02)	0.71 (0.01)	0.70 (0.03)	0.71 (0.02)	**0.73 (0.00)**	**0.73 (0.01)**	**0.73 (0.01)**	**0.73 (0.00)**	**0.73 (0.00)**	**0.73 (0.00)**	0.73 (0.01)	**0.73 (0.00)**
	GM	30	0.93 (0.04)	7.33 (0.19)	6.81 (0.27)	7.08 (0.14)	1.25 (0.03)	1.25 (0.04)	1.26 (0.04)	1.06 (0.05)	1.07 (0.08)	1.06 (0.04)	1.17 (0.20)	1.10 (0.21)	1.07 (0.05)	1.29 (0.07)	**0.96 (0.05)**
		40		8.72 (0.23)	8.74 (0.29)	8.77 (0.27)	1.26 (0.04)	1.24 (0.05)	1.25 (0.06)	1.16 (0.03)	1.12 (0.06)	1.13 (0.05)	1.04 (0.03)	1.16 (0.06)	1.04 (0.03)	1.22 (0.05)	**0.97 (0.05)**
Airfoil	MSE	30	23.17 (1.75)	40.18 (3.45)	60.30 (4.79)	59.22 (4.03)	44.64 (5.63)	63.92 (8.93)	58.96 (9.98)	23.95 (2.42)	23.68 (1.80)	23.86 (2.26)	24.01 (2.44)	23.99 (2.11)	24.01 (2.44)	23.71 (1.97)	**23.46 (2.08)**
		40		59.07 (3.84)	121.62 (22.47)	122.79 (24.81)	58.30 (8.16)	72.40 (24.28)	68.87 (20.67)	28.02 (0.79)	25.28 (1.52)	25.81 (1.86)	28.23 (0.80)	28.08 (0.48)	28.23 (0.80)	25.48 (1.70)	**24.61 (1.60)**
	MAE	30	3.76 (0.22)	4.91 (0.29)	6.21 (0.31)	6.16 (0.27)	5.36 (0.17)	6.55 (0.53)	6.23 (0.57)	3.84 (0.27)	3.79 (0.21)	3.84 (0.24)	3.85 (0.28)	3.86 (0.24)	3.85 (0.28)	3.78 (0.22)	**3.74 (0.24)**
		40		6.37 (0.28)	9.15 (0.98)	9.16 (1.07)	6.15 (0.53)	6.80 (1.39)	6.55 (1.13)	4.27 (0.10)	4.00 (0.14)	4.03 (0.18)	4.30 (0.10)	4.30 (0.08)	4.30 (0.10)	4.01 (0.17)	**3.86 (0.19)**
	Pearson	30	0.72 (0.04)	0.71 (0.04)	0.71 (0.04)	0.71 (0.04)	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	**0.72 (0.04)**	0.72 (0.04)
		40		0.72 (0.03)	0.71 (0.04)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	0.72 (0.03)	**0.73 (0.03)**
	GM	30	2.50 (0.16)	3.17 (0.45)	4.14 (0.37)	4.08 (0.39)	3.64 (0.15)	5.23 (1.19)	4.53 (0.51)	2.52 (0.28)	**2.44 (0.21)**	2.47 (0.21)	2.54 (0.29)	2.47 (0.25)	2.53 (0.30)	2.47 (0.24)	**2.37 (0.21)**
		40		4.44 (0.30)	6.25 (0.80)	6.28 (0.82)	4.02 (0.50)	4.50 (1.05)	4.20 (0.73)	2.97 (0.08)	2.76 (0.18)	2.69 (0.22)	3.00 (0.13)	3.03 (0.13)	3.05 (0.22)	2.75 (0.19)	**2.54 (0.21)**
Auto-mpg	MSE	30	11.11 (1.29)	93.36 (5.38)	45.46 (12.61)	45.90 (9.43)	67.14 (25.25)	30.94 (11.27)	26.01 (4.50)	13.02 (1.85)	12.46 (1.94)	12.50 (2.09)	12.68 (1.95)	14.56 (1.77)	12.84 (1.92)	12.63 (2.09)	**11.54 (1.48)**
		40		153.90 (23.12)	85.61 (17.01)	94.23 (24.25)	116.87 (30.73)	45.75 (21.05)	40.00 (21.19)	21.57 (5.47)	18.20 (5.34)	19.07 (5.41)	20.67 (6.31)	23.98 (6.07)	21.04 (6.22)	18.22 (4.94)	**12.39 (2.70)**
	MAE	30	2.44 (0.15)	8.87 (0.24)	5.23 (1.04)	5.27 (0.65)	7.65 (1.89)	4.70 (1.15)	4.13 (0.49)	2.67 (0.25)	2.57 (0.28)	2.52 (0.30)	2.55 (0.26)	2.87 (0.27)	2.57 (0.28)	2.56 (0.28)	**2.43 (0.21)**
		40		11.69 (1.19)	7.18 (0.77)	7.88 (1.30)	11.06 (2.18)	5.79 (1.80)	5.25 (1.66)	3.64 (0.45)	2.99 (0.38)	3.10 (0.42)	3.29 (0.47)	3.83 (0.49)	3.27 (0.48)	2.98 (0.33)	**2.48 (0.33)**
	Pearson	30	0.91 (0.01)	0.89 (0.02)	0.89 (0.02)	0.89 (0.02)	0.90 (0.01)	0.89 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	**0.91 (0.01)**	0.90 (0.01)	**0.91 (0.01)**	0.90 (0.01)	0.91 (0.02)
		40		0.89 (0.02)	0.90 (0.02)	0.89 (0.02)	0.89 (0.04)	0.88 (0.04)	0.88 (0.04)	0.90 (0.03)	0.90 (0.02)	0.90 (0.01)	0.90 (0.03)	0.89 (0.04)	0.90 (0.03)	0.90 (0.02)	**0.91 (0.02)**
	GM	30	1.59 (0.19)	7.33 (0.72)	3.63 (1.42)	3.51 (0.67)	6.36 (1.57)	3.46 (0.92)	2.91 (0.66)	1.89 (0.33)	1.61 (0.30)	1.74 (0.15)	1.66 (0.22)	1.91 (0.29)	1.60 (0.18)	1.63 (0.31)	**1.56 (0.25)**
		40		8.55 (1.38)	4.58 (1.08)	6.15 (2.23)	8.79 (1.27)	4.58 (1.73)	4.71 (2.53)	2.67 (0.38)	1.72 (0.21)	1.85 (0.18)	2.02 (0.22)	2.84 (0.54)	2.00 (0.71)	1.71 (0.11)	**1.59 (0.21)**
Housing	MSE	30	24.11 (7.02)	42.83 (13.89)	64.42 (24.29)	57.15 (14.15)	45.79 (5.79)	50.48 (14.49)	40.35 (9.13)	27.28 (7.11)	26.27 (6.78)	27.29 (6.89)	27.38 (6.98)	27.35 (7.73)	26.80 (7.48)	26.50 (6.48)	**25.83 (7.12)**
		40		70.36 (12.79)	109.98 (16.74)	91.33 (22.84)	89.01 (8.30)	84.52 (9.19)	84.94 (9.24)	31.64 (5.54)	31.16 (6.31)	30.93 (5.87)	31.70 (5.71)	32.64 (5.39)	31.44 (5.65)	32.05 (6.30)	**28.39 (5.47)**
	MAE	30	3.21 (0.41)	4.79 (0.71)	5.99 (1.38)	5.66 (0.77)	5.69 (0.78)	5.82 (0.64)	5.04 (0.72)	3.59 (0.48)	3.50 (0.36)	3.49 (0.44)	3.55 (0.50)	3.58 (0.53)	3.56 (0.49)	3.45 (0.35)	**3.43 (0.47)**
		40		6.39 (1.05)	7.71 (0.93)	7.17 (1.21)	6.27 (0.38)	6.52 (0.38)	6.54 (0.35)	3.75 (0.27)	3.79 (0.25)	3.72 (0.24)	3.74 (0.25)	3.82 (0.30)	3.72 (0.28)	3.87 (0.20)	**3.58 (0.39)**
	Pearson	30	0.85 (0.03)	0.84 (0.04)	0.84 (0.04)	0.84 (0.04)	0.84 (0.04)	0.83 (0.03)	0.83 (0.04)	0.84 (0.04)	0.83 (0.04)	0.84 (0.04)	0.84 (0.04)	0.84 (0.04)	0.84 (0.04)	0.84 (0.03)	**0.85 (0.03)**
		40		**0.83 (0.05)**	0.80 (0.08)	0.81 (0.07)	0.81 (0.04)	0.80 (0.03)	0.81 (0.04)	**0.83 (0.05)**	0.82 (0.05)	**0.83 (0.05)**	**0.83 (0.05)**	**0.83 (0.04)**	**0.83 (0.05)**	0.82 (0.05)	**0.83 (0.04)**
	GM	30	1.83 (0.22)	2.79 (0.48)	3.90 (0.96)	3.71 (0.64)	3.72 (0.31)	3.77 (0.48)	3.33 (0.38)	2.31 (0.27)	2.15 (0.24)	2.16 (0.24)	2.25 (0.29)	2.35 (0.32)	2.26 (0.22)	2.17 (0.31)	**2.01 (0.22)**
		40		3.91 (0.70)	4.35 (0.72)	4.52 (0.46)	3.42 (0.34)	3.72 (0.60)	3.78 (0.42)	2.48 (0.40)	2.36 (0.20)	2.30 (0.49)	2.32 (0.48)	2.51 (0.54)	2.46 (0.36)	2.31 (0.29)	**2.08 (0.23)**
Concrete	MSE	70	115.78 (12.97)	287.59 (17.80)	290.36 (18.25)	291.17 (18.48)	327.72 (25.02)	309.91 (22.77)	309.93 (22.79)	129.31 (15.44)	124.56 (15.44)	125.82 (16.30)	128.68 (11.69)	128.42 (13.31)	128.28 (11.68)	124.22 (14.47)	**119.87 (11.85)**
		80		296.92 (20.90)	302.81 (23.17)	303.02 (23.73)	327.88 (22.98)	312.69 (22.25)	312.71 (22.26)	131.52 (6.68)	123.90 (8.73)	126.83 (6.99)	130.45 (6.55)	129.06 (8.81)	130.22 (5.96)	122.19 (8.99)	**118.43 (8.06)**
	MAE	70	8.50 (0.58)	13.70 (0.55)	13.75 (0.56)	13.76 (0.56)	14.49 (0.53)	14.13 (0.52)	14.13 (0.53)	9.21 (0.63)	8.87 (0.73)	8.94 (0.73)	9.18 (0.59)	9.17 (0.64)	9.17 (0.59)	8.91 (0.69)	**8.76 (0.64)**
		80		13.84 (0.33)	13.89 (0.33)	13.89 (0.34)	14.36 (0.24)	14.08 (0.28)	14.08 (0.28)	9.30 (0.44)	8.95 (0.41)	9.10 (0.42)	9.24 (0.40)	9.20 (0.50)	9.22 (0.40)	8.97 (0.45)	**8.57 (0.39)**
	Pearson	70	0.78 (0.04)	0.76 (0.05)	0.76 (0.04)	0.76 (0.04)	0.75 (0.05)	0.74 (0.07)	0.74 (0.07)	0.76 (0.04)	0.76 (0.04)	0.76 (0.04)	0.76 (0.04)	0.76 (0.04)	0.76 (0.04)	0.76 (0.04)	**0.77 (0.04)**
		80		**0.78 (0.02)**	0.76 (0.03)	0.77 (0.02)	0.77 (0.03)	0.76 (0.03)	0.77 (0.03)	**0.78 (0.03)**	**0.78 (0.03)**	**0.78 (0.03)**	**0.78 (0.03)**	**0.78 (0.03)**	**0.78 (0.03)**	**0.78 (0.03)**	0.78 (0.03)
	GM	70	5.52 (0.42)	9.01 (0.62)	9.23 (0.55)	9.19 (0.55)	10.10 (0.41)	9.55 (0.46)	9.55 (0.46)	6.24 (0.72)	5.78 (0.77)	5.81 (0.85)	6.22 (0.65)	6.22 (0.70)	6.21 (0.67)	5.84 (0.95)	**5.58 (0.64)**
		80		9.49 (0.35)	9.47 (0.54)	9.12 (0.30)	9.74 (0.17)	9.45 (0.31)	9.45 (0.30)	6.15 (0.46)	6.15 (0.63)	6.05 (0.74)	6.13 (0.54)	6.16 (0.63)	6.14 (0.55)	6.15 (0.61)	**5.54 (0.21)**
Power-plant	MSE	60	25.18 (0.75)	224.57 (49.55)	108.40 (43.46)	190.98 (83.33)	293.95 (52.38)	172.70 (106.10)	194.29 (111.80)	34.76 (1.43)	30.56 (1.55)	32.36 (3.60)	34.08 (1.27)	32.89 (1.62)	34.71 (1.44)	30.87 (1.57)	**25.56 (0.90)**
		70		265.96 (11.44)	190.35 (118.97)	299.00 (8.63)	288.43 (5.14)	171.97 (136.70)	246.64 (150.10)	50.06 (1.84)	43.05 (2.09)	45.07 (0.98)	49.67 (0.98)	46.28 (2.63)	50.23 (2.75)	43.48 (1.39)	**27.25 (0.70)**
	MAE	60	4.00 (0.05)	12.83 (1.23)	8.65 (2.39)	11.65 (2.86)	14.56 (0.12)	11.23 (3.94)	11.92 (4.20)	4.72 (0.14)	4.43 (0.14)	4.53 (0.21)	4.68 (0.10)	4.57 (0.10)	4.74 (0.15)	4.44 (0.11)	**4.05 (0.07)**
		70		13.79 (0.19)	11.29 (4.27)	14.41 (6.00)	14.57 (0.13)	10.80 (4.77)	13.54 (5.12)	5.68 (0.11)	5.23 (0.14)	5.34 (0.16)	5.63 (0.09)	5.40 (0.15)	5.69 (0.06)	5.24 (0.11)	**4.16 (0.07)**
	Pearson	60	0.96 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	**0.96 (0.00)**
		70		**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	0.94 (0.00)	0.94 (0.00)	0.94 (0.00)	**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	**0.95 (0.00)**	0.95 (0.00)
	GM	60	2.76 (0.09)	9.57 (0.48)	6.37 (2.27)	8.18 (2.14)	9.97 (0.18)	8.77 (3.11)	9.43 (3.50)	3.27 (0.13)	3.15 (0.18)	3.11 (0.11)	3.28 (0.21)	3.28 (0.14)	**2.78 (1.44)**	3.16 (0.14)	2.79 (0.07)
		70		9.69 (0.16)	8.59 (4.53)	10.51 (0.28)	11.74 (3.59)	9.41 (4.35)	10.48 (4.19)	3.84 (0.15)	3.53 (0.07)	3.56 (0.14)	3.94 (0.41)	4.39 (0.74)	4.26 (0.60)	3.51 (0.12)	**3.06 (0.21)**