A APPENDICES

A.1 PROOF OF THEOREM 2

Proof. Let's try expanding $\frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \epsilon^2}$ for the first time:

$$\begin{aligned}\frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \epsilon^2} &= \frac{\partial(\frac{\partial L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \epsilon})}{\partial \epsilon} \\ &= \frac{\partial((\frac{\partial L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}})^T \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon})}{\partial \epsilon} \\ &= (\frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta} \partial \epsilon})^T \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon} + (\frac{\partial L(\hat{\boldsymbol{\theta}}_{\epsilon},\mathcal{D})}{\partial \boldsymbol{\theta}})^T \cdot \frac{\partial^2 \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon^2} \\ &= (\frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon})^T \cdot \frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}^2} \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon} + (\frac{\partial L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}})^T \cdot \frac{\partial^2 \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon^2}\end{aligned} \tag{20}$$

To go on, we need to expand $\frac{\partial^2 \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon^2}$, which is as follows:

$$\begin{aligned}\frac{\partial^2 \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon^2} &= \frac{\partial(\frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon})}{\partial \epsilon} \\ &= \frac{\partial(-W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}))}{\partial \epsilon} \\ &= -\frac{\partial(W_{\tilde{\epsilon}}^{-1})}{\partial \epsilon} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}) - W_{\tilde{\epsilon}}^{-1} \cdot \frac{\partial(\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}})}{\partial \epsilon} \\ &= W_{\tilde{\epsilon}}^{-1} \cdot \frac{\partial W_{\tilde{\epsilon}}}{\partial \epsilon} \cdot W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}) - W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n} \cdot \frac{\partial \mathbf{b}'_{\tilde{\epsilon}}}{\partial \epsilon} + \frac{\delta'_{\tilde{\epsilon}}}{\partial \epsilon} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon}) \\ &= W_{\tilde{\epsilon}}^{-1} \cdot (\frac{\partial^3 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}^3} + \delta'_{\tilde{\epsilon}} \cdot \boldsymbol{I}) \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon} - W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n} \cdot \frac{\partial \mathbf{b}'_{\tilde{\epsilon}}}{\partial \epsilon} + \frac{\delta'_{\tilde{\epsilon}}}{\partial \epsilon} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \frac{\partial \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}{\partial \epsilon})\end{aligned} \tag{21}$$

Finally, let $\mathbf{b}''_{\tilde{\epsilon}} = \frac{\partial \mathbf{b}'_{\tilde{\epsilon}}}{\partial \epsilon}$, $\delta''_{\tilde{\epsilon}} = \frac{\partial \delta'_{\tilde{\epsilon}}}{\partial \epsilon}$, $H_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}} = \frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}^2}$ and $T_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}} = \frac{\partial^3 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \boldsymbol{\theta}^3}$

$$\begin{aligned}\frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D})}{\partial \epsilon^2} &= (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}})^T \cdot W_{\tilde{\epsilon}}^{-1} \cdot H_{\hat{\boldsymbol{\theta}}_{\epsilon}} \cdot W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}) \\ &\quad - (\nabla_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D}))^T \cdot W_{\tilde{\epsilon}}^{-1} \cdot (T_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}} + \delta'_{\tilde{\epsilon}} \cdot \boldsymbol{I}) \cdot W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}) \\ &\quad - (\nabla_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D}))^T \cdot W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n} \cdot \mathbf{b}''_{\tilde{\epsilon}} + \delta''_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}) \\ &\quad + (\nabla_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}},\mathcal{D}))^T \cdot W_{\tilde{\epsilon}}^{-1} \cdot \delta'_{\tilde{\epsilon}} \cdot W_{\tilde{\epsilon}}^{-1} \cdot (\frac{1}{n}\mathbf{b}'_{\tilde{\epsilon}} + \delta'_{\tilde{\epsilon}} \cdot \hat{\boldsymbol{\theta}}_{\tilde{\epsilon}})\end{aligned}, \tag{22}$$

which concludes the proof. □

A.2 PROOF OF THEOREM 3

Proof. According to the assumptions in Theroem 3, the error bound depends on $\mathbf{b}'_{\tilde{\epsilon}}, \mathbf{b}''_{\tilde{\epsilon}}, \delta_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}, \delta'_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}$ and $\delta''_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}$.

We first bound $\mathbf{b}'_{\tilde{\epsilon}}$ and $\mathbf{b}''_{\tilde{\epsilon}}$. Note that β_{ϵ} is defined as 4 and $\mathbf{b}'_{\tilde{\epsilon}}$ can be calculated by 13, it is easy to figure out the following: $||\mathbf{b}'_{\epsilon}|| = \mathcal{O}(\frac{1}{\epsilon^2}), ||\mathbf{b}''_{\epsilon}|| = \mathcal{O}(\frac{1}{\epsilon^3})$.

Next, we will bound $\delta_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}, \delta'_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}$ and $\delta''_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}$. With the fact that $\epsilon < 1$, we have $e^{\frac{\epsilon}{4}} - 1 = \frac{\epsilon}{4} + o(\epsilon)$. With the definition of $\delta_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}$ in 4, we have $||\delta_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}|| = \mathcal{O}(\frac{1}{n\epsilon}), ||\delta'_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}|| = \mathcal{O}(\frac{1}{n\epsilon^2})$ and $||\delta''_{\hat{\boldsymbol{\theta}}_{\tilde{\epsilon}}}|| = \mathcal{O}(\frac{1}{n\epsilon^3})$.

Finally, we get our error bound as follows:

$$\mathcal{O}\left(\frac{(\epsilon' - \epsilon)^2}{n^2\epsilon^4}\right),$$

which concludes the proof. □

A.3 SUPPLEMENTARY EXPERIMENTAL RESULTS

Table 2: Average error of estimated loss for measuring point from $\epsilon = 0.05$ to $\epsilon = 0.35$.

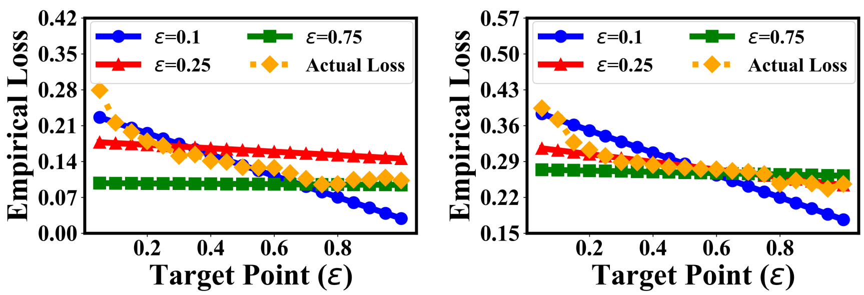
ϵ	0.05	0.10	0.15	0.20	0.25	0.30	0.35
Adlut-LR	0.01982	0.00823	0.00908	0.00408	**0.00003**	0.00033	0.00004
Adlut-SVM	0.08188	0.02052	0.00327	0.00315	0.00445	0.00367	0.00385
Kddcup-LR	0.05884	0.02393	0.00930	0.00721	0.00763	0.00845	0.00960
Kddcup-SVM	0.04735	0.00866	0.00798	0.00793	0.00851	0.01052	0.01143
Gisette-LR	0.18427	0.01482	0.01181	0.02125	0.01931	**0.00302**	0.00810
Gisette-SVM	0.16245	0.00571	0.02480	0.00822	0.00567	0.01023	**0.00550**

Table 3: Average error of estimated loss for measuring point from $\epsilon = 0.40$ to $\epsilon = 0.70$.

ϵ	0.40	0.45	0.50	0.55	0.60	0.65	0.70
Adlut-LR	0.00037	0.00293	0.00309	0.00283	0.00339	0.00623	0.00643
Adlut-SVM	0.00558	**0.00298**	0.00585	0.00890	0.00833	0.00711	0.00546
Kddcup-LR	0.00677	**0.00283**	0.00778	0.00442	0.00768	0.01010	0.01274
Kddcup-SVM	0.00954	0.00848	0.00967	0.00486	0.00363	0.00154	0.00200
Gisette-LR	0.00372	0.00394	0.01361	0.01420	0.01423	0.02461	0.03560
Gisette-SVM	0.00712	0.00823	0.00823	0.00961	0.00926	0.01188	0.01327

Table 4: Average error of estimated loss for measuring point from $\epsilon = 0.75$ to $\epsilon = 1$.

ϵ	0.75	0.80	0.85	0.90	0.95	1.00
Adlut-LR	0.00649	0.01169	0.01480	0.01485	0.01675	0.02276
Adlut-SVM	0.00770	0.00780	0.00840	0.01142	0.01053	0.01020
Kddcup-LR	0.01186	0.01288	0.01200	0.01149	0.01147	0.01213
Kddcup-SVM	0.00112	0.00137	**0.00038**	0.00134	0.00192	0.00250
Gisette-LR	0.04637	0.04476	0.03693	0.03683	0.03308	0.03828
Gisette-SVM	0.01745	0.03641	0.02911	0.03548	0.04540	0.03603



(a) Gisette-LR

(b) Gisette-SVM

Figure 3: Performance of our approximation approach on Gisette with logistic regression (LR) loss and Huber SVM (SVM) loss