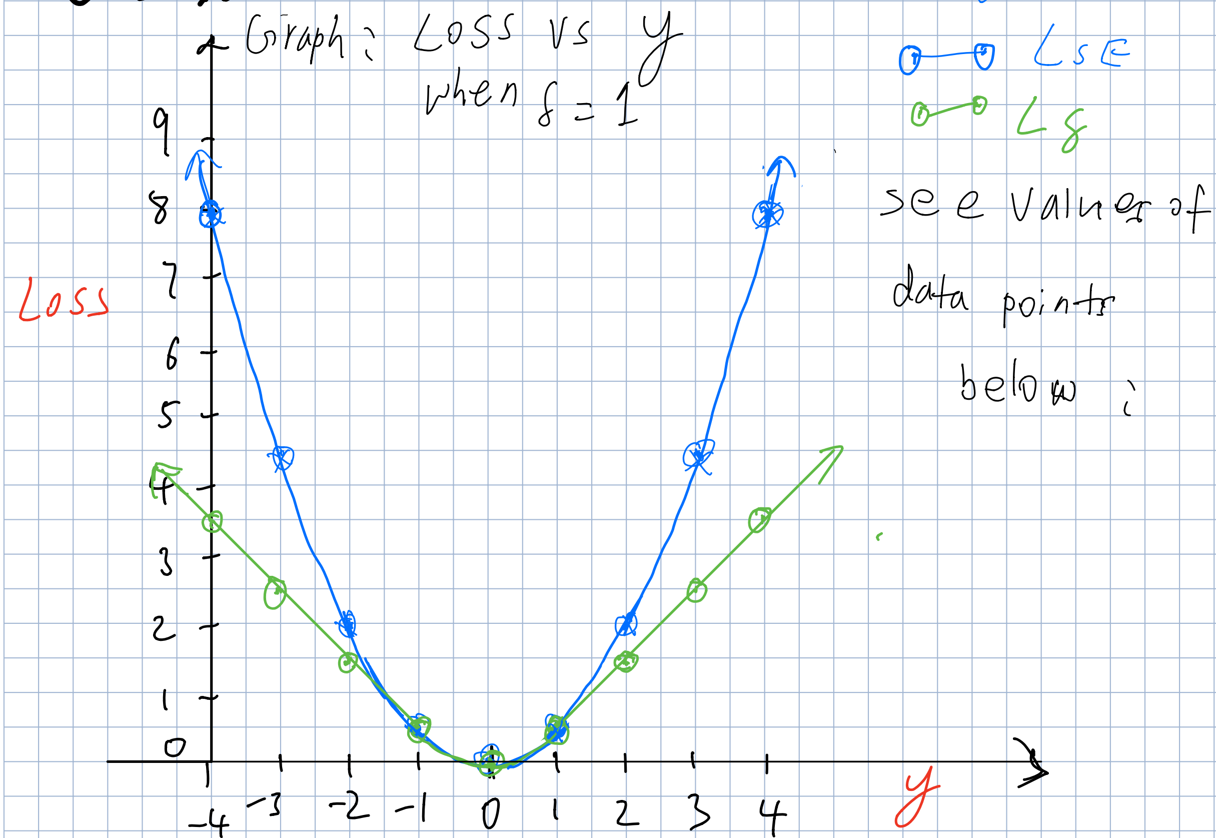
Assignment 2

Q1 a)



see values of data points below:

Values calculated:

$L_{SE}(y, t=0) = \frac{1}{2}y^2$

y	L_{SE}
-4	8
-3	4.5
-2	2

Assume $\delta = 1$

$$L_\delta(y,t) = \begin{cases} \frac{1}{2}y^2 & \text{if } |y| \leq 1 \\ |y| - \frac{1}{2} & \text{if } |y| > 1 \end{cases}$$

$t = 0$

$\delta = 1$

y	L_δ
-4	3.5
-3	2.5

−1	0.5
0	0
1	0.5
2	2
3	4.5
4	8

−2	1.5
−1	0.5
0	0
1	0.5
2	1.5
3	2.5
4	3.5

(Q1b)

Given :

$$H_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } |a| \leq \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{if } |a| > \delta \end{cases}$$

$\frac{1}{2}$

---

1) When $|a| \leq \delta$: $H_\delta(a) = \frac{1}{2}a^2$

$$H'_\delta(a) = \cancel{2} \cdot \cancel{\tfrac{1}{2}}\, a = a \quad 1$$

$\frac{1}{2}\delta$

2) When a is +ve & $a > \delta$:

$$H_\delta(a) = \delta(|a| - \tfrac{1}{2}\delta) = \delta|a| - \delta^2$$

$$H'_\delta(a) = \delta \quad 2$$

3) When a is −ve & $a < -\delta$:

$$H_\delta(a) = \delta(|a| - \tfrac{1}{2}\delta) = \delta|a| - \quad {}^2$$

$$H'_\delta(a) = -\delta \quad 3$$

Thus $H'_\delta(a) = \begin{cases} a & \text{if } |a| \leq \delta \\ \delta & \text{if } a > \delta \\ -\delta & \text{if } a < -\delta \end{cases}$

substitute when $a = y - t$.

$$H'_\delta(y-t) = \begin{cases} y-t & \text{if } |y-t| \leq \delta \\ \delta & \text{if } (y-t) > \delta \\ -\delta & \text{if } (y-t) < -\delta \end{cases} \quad (4)$$

$$\frac{\partial L_\delta}{\partial w} = \frac{\partial H_\delta(y-t)}{\partial w} = \frac{\partial H_\delta(y-t)}{\partial (y-t)} \cdot \frac{\partial (y-t)}{\partial w}$$

$$\frac{\partial L_\delta}{\partial w} = \underbrace{H'_\delta(y-t)}_{\text{sub in } (4)} \cdot \frac{\partial (w^T x + \cancel{b} - \cancel{t})}{\partial w}$$

$$\boxed{\begin{gathered} \text{Answer: } \quad \frac{\partial L_\delta}{\partial w} = H'_\delta(y-t) \cdot x \\[2em] \frac{\partial L_\delta}{\partial w} = \begin{cases} (y-t)x & \text{if } |y-t| \leq \delta \\ \delta x & \text{if } (y-t) > \delta \\ -\delta x & \text{if } (y-t) < -\delta \end{cases} \end{gathered}}$$

$$\frac{\partial L_\delta}{\partial b} = \frac{\partial H_\delta(y-t)}{\partial b} = \frac{\partial H_\delta(y-t)}{\partial (y-t)} \cdot \frac{\partial (y-t)}{\partial b}$$

$$= H'_\delta(y-t) \cdot \frac{\partial (\cancel{w^T x} + \cancel{b} - \cancel{t})}{\cancel{\partial b}}$$

Answer: $\frac{\partial L_\delta}{\partial b} = H'_\delta(y-t)$

$$\frac{\partial L_\delta}{\partial b} = \begin{cases} (y-t) & \text{if } |y-t| \le \delta \\ \delta & \text{if } (y-t) > \delta \\ -\delta & \text{if } (y-t) < -\delta \end{cases}$$

(Q2a)

Let X, Y be the vectors of $\sum x^i, \sum y^i$
A be the vector of $\sum a^i$;

$$J = \frac{1}{2}\sum_{i=1}^{N} a^{(i)}(y^i - w^T x^i)^2 + \frac{\lambda}{2}\|w\|^2$$

convert to matrix form: $J = \frac{1}{2}(Y - Xw)^T A (Y - Xw) + \frac{\lambda}{2}\|w\|^2$

using rule $\|A\| = A \cdot A^T$: $J = \underbrace{\frac{1}{2}(Y - Xw)^T A (Y - Xw)}_{1} + \underbrace{\frac{\lambda}{2} w^T w}_{2}$

Note: We can find w^* by finding w^* where $\frac{\partial J}{\partial w} = 0$

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial (Y - Xw)} \cdot \frac{\partial (Y - Xw)}{\partial w}$$ applied to 1 above

$$\frac{\partial J}{\partial w} = \frac{\partial \left[\frac{1}{2}(Y - Xw)^T A (Y - Xw)\right]}{\partial (Y - Xw)} \cdot \frac{\partial (Y - Xw)}{\partial w} + \frac{\partial (\frac{\lambda}{2} w^T w)}{\partial w}$$

$$\frac{\partial J}{\partial w} = (Y - Xw)^T A \cdot (-X) + \lambda w^T$$

set $\frac{\partial J}{\partial w}$ to 0 to find w^*:

$$(Y - Xw)^T A (-X) + \lambda w^T = 0$$

$$-Y^T A X + w^T X^T A X + \lambda w^T = 0$$

$$-Y^TAX + W^T(X^TAX + \lambda I) = 0$$ transpose all

since A is diagnol matrix $\therefore A^T = A$

$$\rightarrow (X^TAX + \lambda I)W - X^TAY = 0$$

$$(X^TAX + \lambda I)W = X^TAY$$

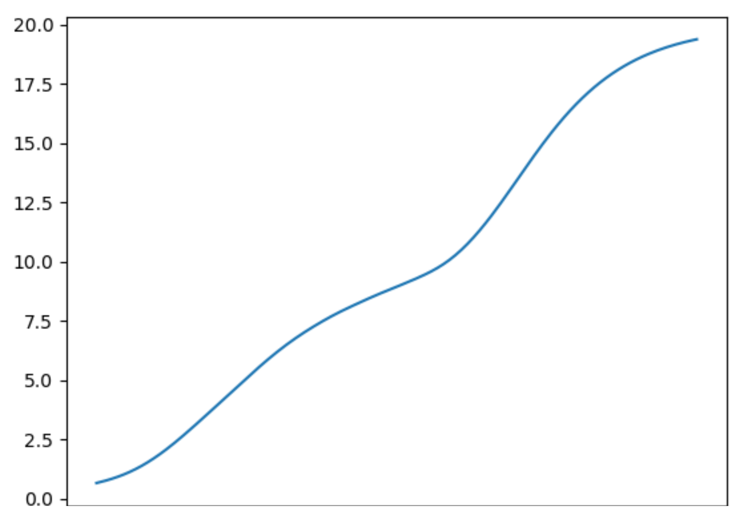
find solution of W: $\boxed{W^* = (X^TAX + \lambda I)^{-1} X^TAY}$

Question 2 c)

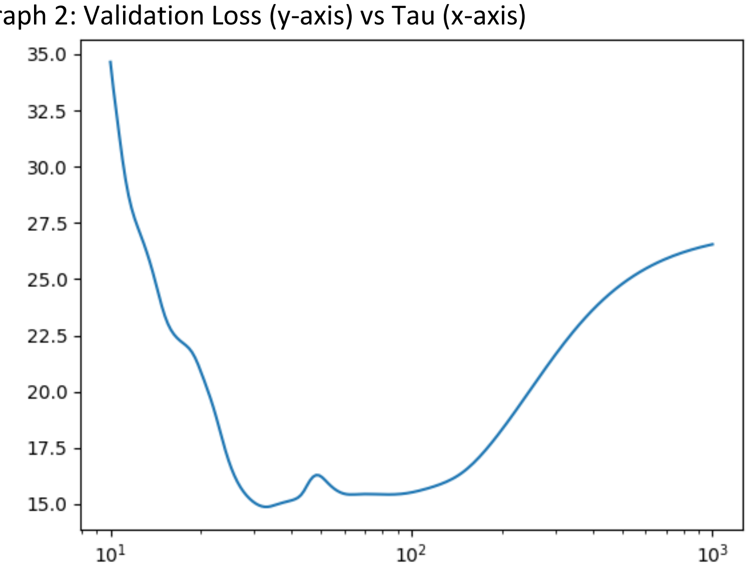
As tau increases, the value of a decreases, which means bigger tau leads to less sensitivity to close neighbors. Thus, I would expect for the training error to increase as tau increases, because the model is giving more significance to the data points further away that likely have different properties than the test point. When tau is small, it's overfitted because a large portion of the weight is contributed by the few training points close to it that are very similar in nature, and so I expect the training loss to be small. For testing error, the situation is a bit more complex as very small tau below a certain limit can lead to overfitting (discussed above) and an increase in testing error. I expect the error to drop as tau increase.

The training error is exactly what I expected. The testing error curve is a bit unexpected as it started to increase again beyond a certain limit and plateaued, which I did not expect. Although that would make sense because the model will likely perform worse by giving too much significance to the far away data points that likely behave very differently than neighboring data points, and results in underfitting. The plateau at the end likely means the model is starting to perform more like a normal linear regression where local weights don't have much more significance than "far-away" weights, and the contribution of weights by different data points become more evenly. The first part of the validation graph was as expected that beyond a below tau value the loss started to increase due to overfitting.

Graph 1: Training Loss (y-axis) vs Tau (x-axis)



Graph 2: Validation Loss (y-axis) vs Tau (x-axis)



Question 3)

Given for t^{th} iteration :

$$h_t \leftarrow \underset{h \in H}{\operatorname{argmin}} \sum_{i=1}^{N} w_i \mathbb{I}\{h(x^i) \neq t^i\},$$

2 sets of i $\begin{cases} E = \{i : h_t(x^i) \neq t^i\} \text{ and } \mathbb{I} = 1 \\ E^c = \{i : h_t(x^i) = t^i\} \text{ and } \mathbb{I} = -1 \end{cases}$ condition 1 from tips.

$$err_t = \frac{\sum_{i=1}^{N} w_i \mathbb{I}\{h_t x^i \neq t^i\}}{\sum_{i=1}^{N} w_i}$$

$$err'_t = \frac{\sum_{i=1}^{N} w'_i \mathbb{I}\{h_t x^i \neq t^i\}}{\sum_{i=1}^{N} w'_i}$$ ← substitute 1 above, bottom split to 2 sets $i \in E^c$ & $i \in E$

$$err'_t = \frac{\sum_{i=1}^{N} w'_i \mathbb{I}\{h_t x^i \neq t^i\}}{\sum_{i \in E} w'_i + \sum_{i \in E^c} w'_i}$$ ← top belong to $i \in E$ according to condition 1

$$err'_t = \frac{\sum_{i \in E} w'_i \cdot (1)}{\sum_{i \in E} w'_i + \sum_{i \in E^c} w'_i}$$

substitute in $w'_i \leftarrow w_i \exp[-\alpha_t t^i h_t(x^i)]$

when $h_t(x^i) \neq t^i$, product of $t^i \cdot h_t(x^i)$ is $-ve$

when $h_t(x^i) = t^i$, product of t^i $h_t(x^i)$ is $+ve$

after substitution :

$$err'_t = \frac{\Sigma_{i \in E} W_i \exp(-\alpha_t \cdot (-1))}{\Sigma_{i \in E} W_i \exp[-\alpha_t \cdot (-1)] + \Sigma_{i \in E^c} W_i \exp(-\alpha_t (1))}$$

$$err'_t = \frac{\Sigma_{i \in E} W_i \exp(\alpha_t)}{\Sigma_{i \in E} W_i \exp(\alpha_t) + \Sigma_{i \in E^c} W_i \exp(-\alpha_t)} \cdot \frac{\exp(\alpha_t)}{\exp(\alpha_t)}$$

Multiply top & bottom of err'_t by $\exp(\alpha_t)$

$$err'_t = \frac{\Sigma_{i \in E} W_i \exp(2\alpha_t) \; (2)}{\underbrace{\Sigma_{i \in E} W_i \exp(2\alpha_t)}_{(2)} + \underbrace{\Sigma_{i \in E^c} W_i}_{(3)}}$$

Note that given $\alpha_t = \frac{1}{2} \log \frac{1 - err_t}{err_t}$

$$2\alpha_t = \log \frac{1 - err_t}{err_t}$$

$$\exp(2\alpha_t) = \frac{1 - err_t}{err_t}$$

(2) : $\Sigma_{i \in E} W_i \exp(2\alpha_t) = \Sigma_{i \in E} W_i \left(\frac{1 - err_t}{err_t}\right)$ (4)

From tip given: $\frac{\Sigma_{i \in E} W_i}{\Sigma_{i=1}^{N} W_i} = err_t$

$$\therefore \frac{1 - err_t}{err_t} = \frac{1 - \frac{\Sigma_{i \in E} W_i}{\Sigma_{i=1}^{N} W_i}}{\frac{\Sigma_{i \in E} W_i}{\Sigma_{i=1}^{N} W_i}} = \frac{\left(\Sigma_{i=1}^{N} W_i - \Sigma_{i \in E} W_i\right) / \cancel{\Sigma_{i=1}^{N} W_i}}{\left(\Sigma_{i \in E} W_i\right) / \cancel{\Sigma_{i=1}^{N} W_i}}$$

cancel out

also note that top part: $\sum_{i=1}^{N} w_i - \sum_{i \in E} w_i = \sum_{i \in E^c} w_i$

So: $\dfrac{1-err_t}{err_t} = \dfrac{\sum_{i \in E_c} w_i}{\sum_{i \in E} w_i}$, substitute into 4

4 is:

$$-\sum_{i \in E} w_i \left(\frac{1-err_t}{err_t}\right) = -\cancel{\sum_{i \in E} w_i} \frac{\sum_{i \in E_c} w_i}{\cancel{\sum_{i \in E} w_i}} = \sum_{i \in E_c} w_i$$

(same as 3)

So

2 = 4 = 3 ∴ 2 = 3

so $\sum_{i \in E} w_i \exp(2\alpha_t) = \sum_{i \in E^c} w_i$

2 = 3

recall above

$$err'_t = \frac{\sum_{i \in E} w_i \exp(2\alpha_t) \;\; 2}{\underbrace{\sum_{i \in E} w_i \exp(2\alpha_t)}_{2} + \underbrace{\sum_{i \in E^c} w_i}_{3}}$$

$$\text{Thus: } err'_t = \frac{\sum_{i \in E_c} w_i}{\sum_{i \in E_c} w_i + \sum_{i \in E_c} w_i} = \frac{1}{2}$$

$$\text{Thus } \boxed{err'_t = \frac{\sum_{i=1}^{N} w'_i \mathbb{I}\{h_t x^i \neq t^i\}}{\sum_{i=1}^{N} w'_i} = \frac{1}{2}}$$