Q4

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

so we want to find w^* argmin $J(w)$

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

$$= \frac{1}{2}(y - Xw)^T A (y - Xw) + \frac{\lambda}{2} w^T w$$

(where y is matrix of $y^{(i)}$, X is matrix of $x^{(i)}$
A is diagnol matrix of $a^{(i)}$)

$$= \frac{1}{2}(y^T A y - y^T A X w - (Xw)^T A y + (Xw)^T A X w + \frac{\lambda}{2} w^T w$$

$$\nabla J(w) = \frac{1}{2}(-2 y^T A X + 2 X^T A X w) + \lambda w$$

$$= -X^T A y + X^T A X w + \lambda w$$

Let $\nabla J(w) = 0$

$$\Rightarrow -X^T A y + X^T A X w^* + \lambda w^* = 0$$

$$(X^T A X + \lambda) w^* = X^T A y$$

$$w^* = (X^T A X + \lambda I)^{-1} X^T A y$$

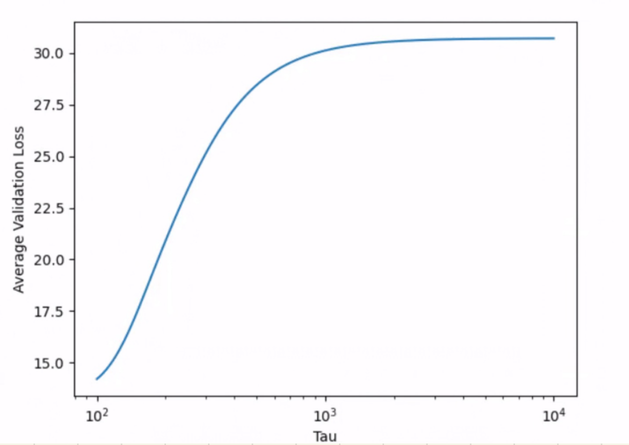
(d) $\tau \to \infty$

then $\dfrac{-\|x-x^{(i)}\|^2}{\tau^2} \to 0$

$\exp\left(\dfrac{-\|x-x^{(i)}\|^2}{\tau^2}\right) \to 1$

$a^{(i)} \to \frac{1}{n}$. which means locally change x doesn't affect. $a^{(i)}$

it become a globally weighted regression



as $\tau > 10^3$

loss doesn't change since any x^i would have similar weights if tau $\to 0$

if x^i is closer to X, a^i is larger than other.

tau $\to \infty$ $\exp\left(\frac{-\|x-x^{(i)}\|^2}{tau^2}\right) \to 0$

if $\|x^i - x\| = 10 \|x^j - x\|$ (x^j is 10 times farther than x^i)

$$a^i / a^j = \left(\frac{1}{\exp(-\|x - x^{(i)}\|^2 / tau^2)} \right)$$ large number

i.e. a^i is weighted much more than a^j

The prediction is over-fit for the x^i near X