Q2.2 Part B

Generalizing to weighted setting

$$J(\theta) = (X\theta - y)^T W (X\theta - y)$$

$$\frac{\partial J}{\partial \theta} = \frac{\partial J}{\partial (X\theta - y)} \cdot \frac{\partial (X\theta - y)}{\partial \theta} = 2\underset{d\times m}{X^T}\, \underset{m\times m}{W} \underset{m\times 1}{(X\theta - y)} = \underset{m\times 1}{0}$$

$$\Rightarrow \quad X^T W X \theta - X^T W y = 0$$

$$\boxed{\theta = (X^T W X)^{-1} (X^T W y)} \quad \#$$

Q2.3 Part(C)

$$w^{(i)} := \exp\left(\frac{-(x - x^{(i)})^T (x - x^{(i)})}{2\tau^2}\right)$$ ← locally weighted linear regression

$$\theta_j' \leftarrow \theta_j - \underset{\text{learning rate}}{\eta} \frac{\partial J}{\partial \theta_j}$$

$$\frac{\partial J}{\partial \theta} = 2X^T (W (X\theta - y))$$ ... from part(B)

[Algo].

for $i = 0 \rightarrow itr$:

$$\theta := \theta - \eta \cdot 2X^T W (X\theta - y)$$

θ: $d\times 1$ vector

X: $m\times d$

W: $m\times m$

y: $m\times 1$

η: 1×1 const.

⇒ locally weighted linear regression is "non-parametric", since parameters θ are computed for every query x

⇒ However, since we already know the closed form solution, instead of G.D. we can simply calculate θ for an input X by $(X^T W X)^{-1} (X^T W y)$