COMP540. HW1

1.1

$X_{m\times d}$. m is number of samples.

$y_{m\times 1}$ d is the dimension

Define $W \triangleq diag(w_1, \dots w_m)$

Let $\underset{m\times 1}{A} = X\theta - y$ $\underset{m\times 1}{A} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$

So, we have $J(\theta) = A^T W A = A^T diag(w_1, \dots, w_m) A$

$$= a_1^2 w_1 + \dots + a_m^2 w_m = \sum_{i=1}^{m} w_i a_i^2$$

$$a_i = \sum_{j=1}^{d} X_{id} \cdot \theta_d - y_i$$

$$= \theta^T \cdot x^{(i)} - y_i$$

So. $J(\theta) = \sum_{i=1}^{m} w_i (\theta^T x^{(i)} - y^{(i)})^2$

$$= \frac{1}{2} \sum_{i=1}^{m} 2w_i (\theta^T x^{(i)} - y^{(i)})^2$$

Let $2w_i = w^{(i)}$. The $W = diag(\frac{w^{(1)}}{2}, \dots \frac{w^{(m)}}{2})$ #

COMP 540 HW1

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

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

$= \theta^T X^T W X \theta - \theta^T X^T W y - y^T W X \theta + y^T W y$

$\frac{\partial J(\theta)}{\partial \theta} = 2 \cdot \theta^T \cdot X^T W X_{*} - y^T W^T X - y^T W X$

$= 2\,\theta^T X^T W X - 2 y^T W X = 0$

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

$X^T W^T X \theta = \cancel{y^T W X} \; X^T W^T y$

$\theta = (X^T W X)^{-1} X^T W y.$

#

COMP 540 HW I

1.3.

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial J}{\partial \theta_j}$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \begin{cases} \frac{1}{2}\sum_{i=1}^{m} w^{(i)} \cdot 2 \cdot \left(\sum_{k=1}^{d}(\theta_k x_k^{(i)}) - y^{(i)}\right) x_j^{(i)} & j \neq 0. \\ \frac{1}{2}\sum_{i=1}^{m} w^{(i)} \cdot 2 \cdot \left(\sum_{k=1}^{d}(\theta_k x_k^{(i)}) - y^{(i)}\right). & j = 0. \end{cases}$$

$$\theta_j \leftarrow \theta_j - \alpha \sum_{i=1}^{m} w^{(i)} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)}\right) x_j^{(i)} \qquad j \neq 0$$

$$\theta_j \leftarrow \theta_j - \alpha \sum_{i=1}^{m} w^{(i)} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)}\right) \qquad j = 0$$

$$x_0^{(i)} = 1$$

input is $\vec{x}$

start with a random $\vec{\theta}$

while (iter < 10000)

$$\theta_j \leftarrow \theta_j - \alpha \sum_{i=1}^{m} w^{(i)} \left(\sum_{k=0}^{d}(\theta_k x_k^{(i)}) - y^{(i)}\right) x_j^{(i)} \qquad j \neq 0$$

$$\theta_j \leftarrow \theta_j - \alpha \sum_{i=1}^{m} w^{(i)} \left(\sum_{k=0}^{d}(\theta_k x_k^{(i)}) - y^{(i)}\right) \qquad j = 0$$

non-parametric, since the # of parametric depends on the # of input data.

COMP 540 HW1

2.1

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(i)} \\ \vdots \\ y^{(m)} \end{pmatrix} \qquad X = \begin{bmatrix} - & x^{(1)} & - \\ - & x^{(2)} & - \\ & \vdots & \\ - & x^{(i)} & - \\ & \vdots & \\ - & x^{(m)} & - \end{bmatrix} \qquad \epsilon = \begin{pmatrix} \epsilon^{(1)} \\ \epsilon^{(2)} \\ \vdots \\ \epsilon^{(i)} \\ \vdots \\ \epsilon^{(m)} \end{pmatrix}$$

$\because$ $\theta = (X^T X)^{-1} X^T y.$ for Least Square Estimator

$y = X\theta^* + \epsilon$

$$\therefore E[\theta] = E[(X^T X)^{-1} X^T y] = E[(X^T X)^{-1} X^T (X\theta^* + \epsilon)]$$

$$= E[(X^T X)^{-1} X^T X \theta^*] + E[(X^T X)^{-1} X^T \epsilon]$$

$$= \theta^* + (X^T X)^{-1} X^T E[\epsilon] = \theta^*$$

#

COMP 540 HW1

2.2

$$Var[\theta] = Var[(X^TX)^{-1}X^T(X\theta^* + \epsilon)]$$

$$= Var[(X^TX)^{-1}X^T\epsilon]$$

$$= (X^TX)^{-1}X^T Var(\epsilon)((X^TX)^{-1}X^T)^T$$

$$= (X^TX)^{-1}X^T \sigma^2 I \quad X(X^TX)^{-1^T}$$

$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= (X^TX)^{-1}\sigma^2$$