Wenxin Xu Hw 2

Problem 1

Define function $f(v) = \sum_{i=1}^{n} |v - x_i|$

$$f'(v) = \sum_{i=1}^{n} \underbrace{\text{sign}(v - x_i)}_{\text{Item}}$$

To set $f'(v) = 0$, the number of positive items should equal the number of negative items, thus $v = \text{median}\{x_1, x_2, \dots, x_n\}$

Answer: $\hat{\mu}$ is the median of the set of x_i.

Problem 2

The likelihood of parameter μ given the observed data x_i is:

$$L(\mu) = \prod_{i=1}^{n} f_\mu(x_i)$$

Log-likelihood is

$$l(\mu) = \log L(\mu) = \sum_{i=1}^{n} \log f_\mu(x_i)$$

$$= \sum_{i=1}^{n} \log \frac{\exp\left(-\frac{(x_i - \mu^*)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

$$= \sum_{i=1}^{n} \frac{-(x_i - \mu^*)^2}{\sqrt{2\pi\sigma^2}\, 2\sigma^2}$$

$$= \sum_{i=1}^{n} \frac{-(x_i - \mu^*)^2}{\sqrt{8\pi\sigma^6}}$$

$$\frac{dl(\mu)}{d\mu} = \sum_{i=1}^{n} \frac{2(\mu^* - x_i)}{\sqrt{8\pi\sigma^6}}$$

set derivative to be 0

$$\sum_{i=1}^{n} \frac{2(\mu^* - x_i)}{\sqrt{8\pi\sigma^6}} = 0$$

$$\sum_{i=1}^{n} (\mu^* - x_i) = 0$$

$$\Rightarrow \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Problem 3

Proof that $e = X\hat{\theta} - y$ is orthogonal to any column of X is to proof that for any vector $v \in span(X)$, $v^T e = 0$

In lecture regression notes we know $\hat{\theta} = (X^T X)^{-1} X^T y$

$$\Rightarrow \quad X^T X \hat{\theta} = X^T y$$

$$X^T X \hat{\theta} - X^T y = 0$$

$$X^T (X\hat{\theta} - y) = 0$$

$$X^T e = 0$$

i.e. $v^T e = 0$

Problem 4

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

where W is the weight matrix

$$W = \begin{bmatrix} d_1 & & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & d_n \end{bmatrix}$$

Proof:

$$\hat{L}_d(\theta) = \frac{1}{n} \sum_{i=1}^{n} d_i (x_i^T \theta - y_i)^2$$

$$\frac{\partial \hat{L}}{\partial \theta_{(j)}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial d_i (x_i^T \theta - y_i)^2}{\partial \theta_{(j)}}$$

$$= \frac{2}{n} \sum_{i=1}^{n} d_i (x_i)_{(j)} (x_i^T \theta - y_i)$$

$$= \frac{2}{n} X^T_{(:j)} W (X\theta - y)$$

$$\Rightarrow \nabla \hat{L}_d(\theta) = \frac{2}{n} X^T W (X\theta - y)$$

set $\nabla \hat{L}_d(\theta)$ to 0

$$\frac{2}{n} X^T W (X\theta - y) = 0$$

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