Jigyas Sharma
3020379

Question 1

The linear model can be written into equation as

$$Y = XB + E \quad \text{where}$$

Y : Observed data $\longrightarrow n\overset{q}{\boxed{}}$

X : Input matrix $\longrightarrow n\overset{p}{\boxed{}}$

B : Parameter Matrix $\longrightarrow p\overset{q}{\boxed{}}$

E : The error matrix $\longrightarrow n\overset{q}{\boxed{}}$

$E = Y - XB$ is the expression that we need to minimize, since smaller errors means better predictions from our linear model.

Let's estimate the parameter matrix, B, by minimizing the Sum of squares of the error (SSE)

$$SSE = \sum_{i=1}^{n} \sum_{k=1}^{q} E[i,k]^2$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{q} (Y - XB)[i,k]^2$$

$$= tr\{(y - XB)^T (y - XB)\}$$

The equation written element wise :

$$f(B) = \sum_{i=1}^{n} \sum_{k=1}^{q} \left(y_{ik} - \sum_{j=1}^{p} x_{ij} B_{jk} \right)^2$$

using $(a-b)^2$

$$f(B) = \sum_{i=1}^{n} \sum_{k=1}^{q} \left(y_{ik}^2 - 2y_{ik} \sum_{j=1}^{p} x_{ij} B_{jk} + \left(\sum_{j=1}^{p} x_{ij} B_{jk} \right)^2 \right)$$

Now, we can differentiate each term using chain rule

$$\frac{d}{dB_{jk}} (y_{ik}^2) = 0$$ since y_{ik}^2 change is independent of B.

$$\frac{d}{dB_{jk}} \left(-2 y_{ik} \sum_{j=1}^{p} x_{ij} B_{jk} \right) = -2 x_{ij} y_{ik}$$

considering simple form $-2y \sum x B$

differentiation $= -2yx$

since differention is 0 when $j \neq k$

$$\frac{d}{dB_{jk}}\left(\sum_{j=1}^{p} x_{ij} B_{jk}\right)^2 = 2x_{ij}\sum_{j=1}^{p} x_{ij} B_{jk}$$

consider $S = \sum_{j=1}^{p} x_{ij} B_{jk}$

then $\frac{d}{dB_{jk}}(S^2) = 2S \cdot \frac{dS}{dB_{jk}}$

$$\frac{dS}{dB_{jk}} = x_{ij}$$

$$\frac{d}{dB_{jk}}(S^2) = 2\left(\sum_{j=1}^{p} x_{ij} B_{jk}\right) \cdot x_{ij}$$

Combining the derivatives

$$\frac{d(f(B))}{dB} = 0 + \sum_{i=1}^{n}\left(-2x_{ij} y_{ik} + 2x_{ij}\sum_{j=1}^{p} x_{ij} B_{jk}\right)$$

taking $2x_{ij}$ as common

$$= 2\sum_{i=1}^{n} x_{ij}\left(\sum_{j=1}^{p} x_{ij} B_{jk} - y\right)$$

Re write in matrix form

$$= 2X^T(XB - y)$$
$$= -2X^T(y - XB)$$

Question 2

2) Lets consider that the random variable Z follows the normal distribution.

$$Z \sim N(\mu, \sigma^2)$$

then the probability density function of z is given by

$$f_z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)}$$

Now,

$$E[i,k] \sim N(0, \sigma^2)$$

and the model has a parameter matrix B and parameter σ^2.

Then to estimate σ^2 we can use the Maximum Likelyhood estimation.

Since, there is an independence in $E[i,k]$ the joint pdf is given by the product of marginal pdfs.

$$L(\hat{B}, \sigma^2) = \prod_{i=1}^{n} \prod_{k=1}^{q} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{e[i,k]^2}{2\sigma^2}\right)}$$

For any fixed σ^2, the above expression is maximized if $\sum_{i=1}^{n} \sum_{k=1}^{q} (y - XB)[i,k]^2$ is minimized.

Since this is a Least square Estimates (LSE) we have the solution

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

This the MLE for B.

Now, maximizing with respect to σ^2 we take the derivative of the log-likelihood with respect to σ^2

$$\frac{d}{d\sigma^2} \log L(\hat{B}, \sigma^2) = \frac{-nq}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} \sum_{k=1}^{q} (y-XB)E(i,k)^2 = 0$$

multiply $2\sigma^4$

$$-nq\sigma^2 + \sum_{i=1}^{n} \sum_{k=1}^{q} (y-XB)E(i,k)^2 = 0$$

solving for σ^2

$$\boxed{\sigma^2 = \frac{1}{nq} \sum_{i=1}^{n} \sum_{k=}^{q} (y-XB)E(i,k)^2}$$

Then the MLE for σ^2 is given by above expression.