Assignment 1 - EE708

Name : Kethan Challa
Roll No : 230317
Branch : CSE

Question 1)

① Expected error can be broken down as

$$\text{error} = (\text{bias})^2 + \text{variance} + \text{noise}$$

→ When the data is small, it is overfitted (seems like)
→ where the model cannot properly learn the pattern of data
→ so this makes high variance
→ So, we need to reduce the variance, so have to use SIMPLE MODELS instead of complex.
→ Because they have → High Bias
→ Low variance
which doesn't have issue with small dataset.

~~Example : For predicting CATS vs DOGS images, with COMPLEX models, It gives importance to COLOUR, BACKGROUND, FUR, etc rather than FACIAL features with some 50 images.~~

~~but a SIMPLE model, it may give importance to them.~~

Example : Predicting stock prices. for 10 days dataset.

Complex Model :
With short data, it gives importance to fluctuations rather REAL TRENDS. → High Variance

SIMPLE MODEL :
Assumes linear (UP/DOWN) trend, which yields better results.

② $E = \sum_t [r^t - g(x^t|w)] + \lambda \sum_i w_i^2$

→ Regularisation term.
→ Used for reducing w values.
→ To reduce overfitting.

λ ⇒ Regularisation Constant

→ On ① Increasing λ

- → Weights Decrease
- → Simple Model
- → Increase Bias
- → Decrease Variance.
- → Underfit

② Decreasing λ (very less)

- → Weights Increase
- → Complex Model
- → Decrease Bias
- → Increase Variance.
- → Overfit.

Assignment - 2

③ Assuming linear regression model, $y_n = w_0 + w_1 x_n$.

$$S_{xy} = \sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}$$

$\bar{x}, \bar{y} \rightarrow$ average of the values x, y over 'N' samples

a) Least square estimate of Slope & Intercept.
(w_1) (w_0)

$$\Rightarrow \boxed{\hat{w}_1 = \frac{S_{xy}}{S_{xx}}} \quad \boxed{\hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}}$$

$$\Rightarrow S_{xy} = 1996904.15 - (250)\left(\frac{11211.00}{250}\right)\left(\frac{44520.80}{250}\right)$$

$$S_{xy} = 449.27$$

$$\Rightarrow S_{xx} = 543503.00 - (250)\left(\frac{11211.00}{250}\right)\left(\frac{11211.00}{250}\right)$$

$$S_{xx} = 40756.916$$

$$\hat{w}_1 = \frac{S_{xy}}{S_{xx}} = 0.011$$

$$\hat{w}_0 = \left(\frac{44520.80}{250}\right) - (0.011)\left(\frac{11211}{250}\right)$$

$$= 177.5899$$

$\therefore$ Slope $\Rightarrow$ 0.011
Intercept $\Rightarrow$ 177.59

$$\Rightarrow y = 0.011x + 177.59$$

(b) 25 year old

$\Rightarrow y = (25)(0.011) + 177.59$

$y = 177.865$ lbs.

(c) Residual = Observed value − Predicted value

$= 170 - 177.865$ lbs.

$= -7.865$ lbs.

(d) ↳ The residual is −ve,
the model overestimated the value of weight
since predicted values was higher than original.

$= -7.865$ lbs.

(d) ↳ The residual is -ve, the model overestimated the value of weight since predicted values was higher than original.

(4).

$N = 14$

(a). $S_{xy} = \sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}$

$\rightarrow S_{xy} = 1697.8 - (14)\left(\frac{572}{14}\right)\left(\frac{43}{14}\right)$

$= 1697.8 - 1756.86$

$S_{xy} = -59.06$

$\rightarrow S_{xx} = (157.42) - (14)\left(\frac{43}{14}\right)\left(\frac{43}{14}\right)$

$= +25.35$

$\Rightarrow \boxed{\hat{w}_1 = \frac{S_{xy}}{S_{xx}} = -2.33}$

$\hat{w}_0 = \bar{y} - \hat{w}_1\bar{x}_0 \Rightarrow \frac{572}{14} + 2.33 \cdot \frac{43}{14}$

$\hat{w}_0 = \frac{672.89}{14} = 48.01$ $\boxed{48.01}$

$\rightarrow y_n = 48.01 - 2.33 x_n$

Variance of Error. $\Rightarrow \sigma^2 = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2$

$$\sigma^2 = \frac{1}{n-2} \left[\sum_i y_i^2 - \hat{w}_0 \sum_i y_i - \hat{w}_1 \sum_i x_i y_i \right]$$

$$\underline{\sigma^2 = 1.84}$$

(b) Based on $y_n = 48.01 - 2.33 x_n$

For $x = 4.3 \rightarrow y = 48.01 - 10.019$

$y = 37.99$

(c) $y_n = 48.01 - 2.33 \times 3.7 = 48.01 - 8.621$

$y = 39.389$.

(d) At $x = 3.7 \rightarrow$ Observed $= 46.1$

Predicted $= 39.389$.

Residual $= 46.1 - 39.389$

$\underline{= 6.711}$

The model underpredicted/underestimated.

Q5) (a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

$$g(y) = \frac{1}{10} \sum_n' ((\beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n}) - y_n)^2$$

$$\frac{\partial g(y)}{\partial \beta_0} = 0, \quad \frac{\partial g(y)}{\partial \beta_1} = 0; \quad \frac{\partial g(y)}{\partial \beta_2} = 0$$

$$\rightarrow \frac{\partial g(y)}{\partial \beta_0} = 0 \Rightarrow \frac{1}{5} \sum' (\beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} - y_n) = 0.$$

$$\Rightarrow \beta_0 \sum' 1 + \beta_1 \sum' x_{1n} + \beta_2 \sum' x_{2n} = \sum' y_n.$$

$$\Rightarrow \quad \rightarrow (1)$$

$$\rightarrow \frac{\partial g(y)}{\partial \beta_1} = 0 \rightarrow \frac{1}{5} \sum' (\beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} - y_n)(x_{1n}) = 0$$

$$\Rightarrow \beta_0 \sum' x_{1n} + \beta_1 \sum' x_{1n}^2 + \beta_2 \sum' x_{1n} x_{2n} = \sum' x_{1n} y_n \rightarrow (2)$$

$$\rightarrow \frac{\partial g(y)}{\partial \beta_2} \Rightarrow \frac{1}{5} \sum' (\quad) = 0$$

which gives $\rightarrow$

$$\beta_0 \sum' x_{2n} + \beta_1 \sum' x_{1n} x_{2n} + \beta_2 \sum' x_{2n}^2 = \sum' x_{2n} y_n.$$

(1), (2), (3) are normal eqns.

(b) Substituting them

$\rightarrow$ ① : $10\beta_0 + 223\beta_1 + 553\beta_2 = 1916$

$\rightarrow$ ② : $223\beta_0 + 5200.9\beta_1 + 12352\beta_2 = 43550.$

$\rightarrow$ ③ : $553\beta_0 + 12352\beta_1 + 31729\beta_2 = 104736.8.$

$\therefore \beta_0 = 171.06$

$\beta_1 = 3.71$

$\beta_2 = -1.73.$

(c) $y_{pred} = (171.06) + (3.71)x_1 - (1.73)x_2 + \epsilon$

$x_1 = 18$

$x_2 = 43\%.$

$y_{predicted} \cong \underline{189.471}$

6) given $(X^T X)^{-1}$ and $X^T Y$

Parameter Matrix = $\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_2 \end{bmatrix}_{3\times 1}$

Dimension is 3 because in given matrix is 3×3 which is possible only where there are 3 parameters

$W = (X^T X)^{-1} X^T Y$ = Multiply both matrices.

$$W = \begin{bmatrix} 2.9705 & -4.0042e-2 & -4.1679e-2 \\ -0.4004 & 6.0774e-4 & -7.3875e-5 \\ -0.00417 & -7.3875e-5 & 2.5766e-4 \end{bmatrix} \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

$$W = \begin{bmatrix} 2.9705 & -0.040042 & -0.041679 \\ -0.4004 & +0.00060774 & -0.000073875 \\ -0.00417 & -0.000073875 & 0.00025766 \end{bmatrix} \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

$$W = \begin{bmatrix} -6744.13 \\ -1715.15 \\ 1.763 \end{bmatrix}$$

$\Rightarrow$ $y = -6744.13 + (-1715.15)x_1 + (1.763)x_2$

Where x_1, x_2 are height, waist.

y is percent of BF

7) We can do the least square estimates in 2 Methods.

① Matrix (Design Matrix, $(X^TX)^{-1}X^TY$)

② Differentiate and Equate $= 0$.

①

$$X = \begin{bmatrix} 1 & (x_1)_1 & (x_2)_1 & (x_1)_1(x_2)_1 & (x_1^2)_1 & (x_2^2)_1 \\ 1 & (x_1)_2 & (x_2)_2 & (x_1)_2(x_2)_2 & (x_1^2)_2 & (x_2^2)_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (x_1)_n & (x_2)_n & (x_1)_n(x_2)_n & (x_1^2)_n & (x_2^2)_n \end{bmatrix}_{n \times 6}$$

$$X^TX = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum (x_1 x_2) & \sum x_1^2 & \sum x_2^2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1^2 x_2 & \sum x_1^3 & \sum x_1 x_2^2 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 & \sum x_1 x_2^2 & \sum x_1^2 x_2 & \sum x_2^3 \\ \sum x_1 x_2 & \sum x_1^2 x_2 & \sum x_1 x_2^2 & \sum x_1^2 x_2^2 & \sum x_1^3 x_2 & \sum x_1 x_2^3 \\ \sum x_1^2 & \sum x_1^3 & \sum x_1^2 x_2 & \sum x_1^3 x_2 & \sum x_1^4 & \sum x_1^2 x_2^2 \\ \sum x_2^2 & \sum x_1 x_2^2 & \sum x_2^3 & \sum x_2^3 & \sum x_1 x_2^3 & \sum x_2^4 \end{bmatrix}$$

Final $\rightarrow$ $W = (X^TX)^{-1} X^T Y$.