Solution of MST UM501 (Machine Learning)

1	Finding means $\bar{X}_1 = \frac{1}{4}(4+8+13+7) = 8,$ $\bar{X}_2 = \frac{1}{4}(11+4+5+14) = 8.5.$ (Mark 1) Finding covariance matrix $S = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) \end{bmatrix}$ $= \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$ (Mark 1) Finding Eigen values $0 = \det(S - \lambda I)$ $= \begin{vmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{vmatrix}$ $= (14-\lambda)(23-\lambda) - (-11) \times (-11)$ $= \lambda^2 - 37\lambda + 201$ $\lambda = \frac{1}{2}(37 \pm \sqrt{565})$ $= 30.3849,\ 6.6151$ $= \lambda_1,\ \lambda_2 \quad \text{(say)}$ (Mark 1) Finding the principle component $e_1 = \begin{bmatrix} 11/\lVert U_1 \rVert \\ (14-\lambda_1)/\lVert U_1 \rVert \end{bmatrix}$ $= \begin{bmatrix} 11/19.7348 \\ (14-30.3849)/19.7348 \end{bmatrix}$ $= \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$ (Mark 1) **Finding the final dimension** -4.3052 3.7361 5.6928 -5.1238 (Mark 1)
2	Regularization: While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it. (1+4 marks)

Ridge Regression using Gradient Descent

- We know in gradient descent optimization, we update the regression coefficients as follows:

$$\beta_j = \beta_j - \alpha \frac{\partial(J(\beta))}{\partial \beta_j} \; for \; j = 0,1,2, \dots \dots .. k$$

- For Ridge Regression, cost function is given by:

$$J = \frac{1}{2n}\sum_{i=1}^{n}((y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \cdots \dots \dots \dots .. - \beta_k x_{ik})^2 + \lambda \sum_{i=0}^{k} \beta_i^2)$$

- The gradient of cost function w.r.t β's is given by:

$$\frac{\partial J}{\partial \beta_0} = \frac{1}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) + \lambda\, \beta_0)$$

Similarly, $\frac{\partial J}{\partial \beta_1} = \frac{1}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{i1} + \lambda\, \beta_1)$

$$\frac{\partial J}{\partial \beta_2} = \frac{1}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{i1} + + \lambda\, \beta_2)$$

$$\frac{\partial J}{\partial \beta_3} = \frac{1}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{i3} + + \lambda\, \beta_3)$$

$$\vdots$$

In general, $\boldsymbol{\frac{\partial J}{\partial \beta_j} = \frac{1}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{ij} + \lambda\, \beta_j)}$

- Therefore, the regression coefficients are updated as:

$$\beta_j = \beta_j - \frac{\alpha}{n}\sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{ij} + \lambda\, \beta_j)$$

or $\boldsymbol{\beta_j = \beta_j\left(1 - \frac{\alpha\lambda}{n}\right) - \frac{\alpha}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} - y_i) \times x_{ij}}$

Since the factor $\left(\mathbf{1} - \frac{\alpha\lambda}{n}\right)$ is less than 1, therefore, the algorithm, will keep on shrinking the values of the regression coefficients, and will handle the problem of overfitting.

3

Q3

(a) part (2.5 marks)

hours = 33 hrs

$y = -64 + 2 \times 33 = 2$ —— 1 mark

probability, $z = \frac{1}{1+e^{-y}} = \frac{1}{1+e^{-2}} = \frac{1}{1+0.135} = 0.88$

1 mark

As $z > 0.5$ so predicted result = 1

given actual result = 1

so predicted result and actual result is same

→ 0.5 mark

(b) part (2.5 marks)

$0.95 < \frac{1}{1+e^{-y}}$

$e^{-y} < \frac{1}{0.95} - 1$

$y = \log_e\left(\frac{0.95}{1-0.95}\right) \Rightarrow$

$y = 2.9443$

2 marks

Hours $= \frac{y+64}{2} = \frac{66.9443}{2} = 33.47$ — 0.5 marks

Ans hours ≥ 33.47 hrs

4

Predicted \ Actual	A	B	C	D
A	25	48	90	70
B	12	14	16	30
C	88	40	17	11
D	33	24	13	18

$Precision_A = \cancel{\frac{25}{25+12+88}} \quad \frac{25}{25+48+90+70} = \frac{25}{233} = 0.107$

$Precision_B = \frac{14}{72} = 0.194$

$Precision_C = \frac{17}{156} = 0.109$

$Precision_D = \frac{18}{88} = 0.2045$

$Sensitivity_A = 0.158 = \frac{25}{158}$

$Sensitivity_B = \frac{14}{126} = 0.111$

$Sensitivity_C = \frac{17}{136} = 0.125$

$Sensitivity_D = \frac{18}{129} = 0.1395$

c 88 /40 17 11 30

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{14+16+30+40+17+11+24+13+18 = 183}{183+208} = 0.468$$

$$\text{False Positive Rate} = 1 - \text{Specificity} = 0.53196$$

There are 10 calculations of 0.5 marks each. In case any student has written correct formula for all 4 parts and nothing else is correct 1 mark is given.

5 **Solution:**

61	68	68	64	65	70	63	62	64	67
112	123	130	115	110	125	100	113	116	125

	X	Y	X2	XY
	61	112	3721	6832
	68	123	4624	8364
	68	130	4624	8840
	64	115	4096	7360
	65	110	4225	7150
	70	125	4900	8750
	63	100	3969	6300
	62	113	3844	7006
	64	116	4096	7424
	67	125	4489	8375
Sum	652	1169	42588	76401
Average	65.2	116.9	4258.8	7640.1

1822
776

B1	2.347938144

B0	-36.18556701

(01 Mark)

$$\beta_1\hat{} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\text{n}\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

(0.5 marks)

B1	2.347938144

(1 marks)

$$\beta_0\hat{} = \overline{y} - \beta_1\hat{}\,\overline{x}$$

(0.5 Marks)

B0	-36.18556701

(1 mark)

(b) **Y= 2.3479 X – 36.1855**
= 2.3479 (69) - 36.1855
= 162.0051 – 36.1855
= 125.8196

(1 mark)