Scheme of Valuation/Answer Key

(Scheme of evaluation (marks in brackets) and answers of problems/key)

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

M.Tech Degree S1 (S,FE) May 2024 (2022 scheme) / S1 (WP) (R) December 2023 Examination

Discipline: COMPUTER SCIENCE AND ENGINEERING

Course Code & Name: 221TCS100 ADVANCED MACHINE LEARNING

Max. Marks: 60 | Duration: 2.5 Hours

PART A

		Answer all questions. Each question carries 5 marks	Marks
	1	Gradient Descent is an iterative optimization algorithm that is used to find the values of parameters of a function that minimizes a cost function. In gradient descent, we need to choose the learning rate, Number of iterations, and another hyperparameter. Gradient descent works well with large number of features. Feature scaling can be used– 2.5 marks Normal Equation is an analytical approach used for optimization. Normal equations directly compute the parameters of the model that minimizes the Sum of the squared difference between the actual term and the predicted term of the dataset without needing to choose any hyperparameters like learning rate or the number of iterations. Normal equation works well with small number of features. No need of feature scaling– 2.5 marks	(5)
	2	Basic components – Input layer, weights, bias, activation function, output – 2 marks. Figure – 1 mark Any 2 activation functions – 2 marks.	(5)
	3	Dimensionality reduction is the process of reducing the number of variables under consideration by obtaining a smaller set of principal variables – 1 mark. PCA finds a new set of k features that are the combination of the original n features. PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is	(5)

		less than or equal to the smaller of the number of original variables or the number of observations. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. – 4 marks Explanation with suitable diagram is also correct.	
	4	F | C | t | t | $0.1 \times 0.8 = 0.08$ t | f | $0.1 \times 0.2 = 0.02$ f | t | $0.9 \times 0.3 = 0.27$ f | f | $0.9 \times 0.7 = 0.63$	(5)
	5	Precision and Recall – 1 mark each. Increase threshold to increase precision – 1.5 marks. Decrease threshold to increase recall – 1.5 marks.	(5)
PART B			
Answer any 5 questions. Each question carries 7 marks			
	6	a) Supervised and unsupervised learning algorithm – 1 mark each. Examples, supervised - Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc. – 0.5 marks for 1 example Examples, unsupervised – Kmeans, hierarchichal clustering, Principal component analysis etc. - 0.5 marks for 1 example b) Maximum Likelihood Estimation (MLE) and Maximum A Posteriori (MAP) estimation are method of estimating parameters of statistical models. While MLE uses the likelihood to estimate the parameters, MAP will consider both the likelihood and prior knowledge of the system state. – 2 marks. Therefore, if the observed data size is small, MAP will give a better estimate – 1 mark.	(4)

7 (7)

$$
\begin{aligned}
n &= 5\\
\bar{x} &= \frac{1}{5}(1.0 + 2.0 + 3.0 + 4.0 + 5.0)\\
&= 3.0\\
\bar{y} &= \frac{1}{5}(1.00 + 2.00 + 1.30 + 3.75 + 2.25)\\
&= 2.06\\
\text{Cov}(x, y) &= \frac{1}{4}[(1.0 - 3.0)(1.00 - 2.06) + \cdots + (5.0 - 3.0)(2.25 - 2.06)]\\
&= 1.0625\\
\text{Var}(x) &= \frac{1}{4}[(1.0 - 3.0)^2 + \cdots + (5.0 - 3.0)^2]\\
&= 2.5\\
b &= \frac{1.0625}{2.5}\\
&= 0.425\\
a &= 2.06 - 0.425 \times 3.0\\
&= 0.785
\end{aligned}
$$

8 ( )

Classes, C_1 – has flu, C_2 – does not have flu.

Frequency Table:

	Chills		Running Nose		Headache			Fever		Total
	Yes	No	Yes	No	No	Mild	Strong	Yes	No	
Flu -Yes	3	2	4	1	1	2	2	4	1	5
Flu - No	1	2	1	2	1	1	1	1	2	3
Total	4	4	5	3	2	3	3	5	5	8

Given symptoms = { chills, fever, mild headache, no running nose}. Let us call this, set S.

Calculate P(S | Flu - Yes) and P (S | Flu – No) and choose the maximum of two.

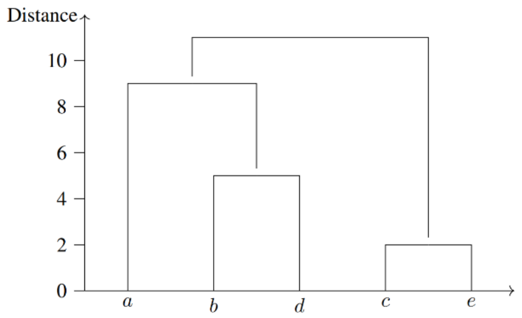
P(S | Flu-Yes) = P(chills | Flu-Yes) * P(fever | Flu-Yes) * P(mild headache | Flu-Yes) * P(no running nose | Flu-Yes)

P (chills | Flu-Yes) = P(Flu-yes | chills) * P (chills) / P(Flu-yes)

= (¾ * 3/8) / 5/8

Similarly, calculate others.

Steps – 6 marks (3 marks for each class). Final answer – 1 mark.

<table>
<tr><td></td><td>9</td><td>Distance Calculation – 4 Marks. Dendrogram construction – 3 marks.

</td><td>(7)</td></tr>
<tr><td></td><td>10</td><td>a) Overfitting – 1 mark. Any cross validation technique – 3 marks.
b) GMM – Unsupervised learning algorithm, GMM does not use a distance measure, but applies a probability distribution around the cluster centers to work out the likelihood that a data point belongs to a given cluster, can analyze more complex and mixed data, can handle outliers more easily. Does not directly assign data points to clusters.</td><td>(4)

(3)</td></tr>
<tr><td></td><td>11</td><td>

$$K(x,y) = (x\cdot y+1)^2 - 1$$
$$= \left((x\cdot y)^2 + 2(x\cdot y)\times 1 + \cancel{1^2}\right) - \cancel{1}$$
$$= (x\cdot y)^2 + 2(x\cdot y) \quad \text{— } (1)$$
$$x\cdot y = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1y_1 + x_2y_2$$
$$2 \qquad (1) = (x_1y_1 + x_2y_2)^2 + 2x_1y_1 + 2x_2y_2$$
$$= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 + 2x_1y_1 + 2x_2y_2$$
$$= \varphi(x)\cdot\varphi(y)$$
where
$$\varphi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}\,x_1x_2 \\ \sqrt{2}\,x_1 \\ \sqrt{2}\,x_2 \end{bmatrix} \qquad \varphi(y) = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}\,y_1y_2 \\ \sqrt{2}\,y_1 \\ \sqrt{2}\,y_2 \end{bmatrix}$$

</td><td>(7)</td></tr>
</table>

	12	Accuracy – 1. Bagging/ Boosting/ Random forest – Any two methods – 3 marks each.	(7)