

SPRING MID SEMESTER ML EVALUATION SCHEME -2025
School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Subject Name: Machine Learning
[Subject Code: CS31002]

1. Answer all the questions.

- a) Given three 2D vectors such as $\mathbf{a} = (2, 5)$, $\mathbf{b} = (-3, 7)$, and $\mathbf{c} = (4, -2)$, find out which two vectors are the closest to each other based on cosine similarity?

Ans: To determine which two vectors are the closest based on cosine similarity, we need to calculate the cosine similarity between each pair of vectors. The cosine similarity between two vectors \mathbf{u} and \mathbf{v} is given by:

$$\text{cosine similarity} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$$

where $\mathbf{u} \cdot \mathbf{v}$ is the dot product of the vectors, and $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ are the magnitudes (norms) of the vectors.

We will compute:

$$\cos(a, b), \cos(b, c), \text{ and } \cos(a, c)$$

The highest cosine similarity indicates the closest pair in terms of direction.

Dot Products:

$$\mathbf{a} \cdot \mathbf{b} = (2)(-3) + (5)(7) = -6 + 35 = 29,$$

$$\mathbf{a} \cdot \mathbf{c} = (2)(4) + (5)(-2) = 8 - 10 = -2,$$

$$\mathbf{b} \cdot \mathbf{c} = (-3)(4) + (7)(-2) = -12 - 14 = -26.$$

Norms:

$$\|\mathbf{a}\| = \sqrt{2^2 + 5^2} = \sqrt{4 + 25} = \sqrt{29} \approx 5.3852,$$

$$\|\mathbf{b}\| = \sqrt{(-3)^2 + 7^2} = \sqrt{9 + 49} = \sqrt{58} \approx 7.6158,$$

$$\|\mathbf{c}\| = \sqrt{4^2 + (-2)^2} = \sqrt{16 + 4} = \sqrt{20} \approx 4.4721.$$

Calculate cosine similarities:

Cosine Similarity between \mathbf{a} and \mathbf{b} :

$$\text{cosine similarity}(\mathbf{a}, \mathbf{b}) = \frac{29}{\sqrt{29} \times \sqrt{58}} = \frac{29}{\sqrt{29 \times 58}} = \frac{29}{\sqrt{1682}} \approx \frac{29}{41} \approx 0.707$$

Cosine Similarity between \mathbf{a} and \mathbf{c} :

$$\text{cosine similarity}(\mathbf{a}, \mathbf{c}) = \frac{-2}{\sqrt{29} \times 2\sqrt{5}} = \frac{-2}{2\sqrt{145}} = \frac{-1}{\sqrt{145}} \approx \frac{-1}{12.04} \approx -0.083$$

Cosine Similarity between \mathbf{b} and \mathbf{c} :

$$\text{cosine similarity}(\mathbf{b}, \mathbf{c}) = \frac{-26}{\sqrt{58} \times 2\sqrt{5}} = \frac{-26}{2\sqrt{290}} = \frac{-13}{\sqrt{290}} \approx \frac{-13}{17.03} \approx -0.763$$

Since, $\cos(a, b) = 0.707$ is the largest compared to -0.083 and -0.763.

- b) Training accuracy is 100% for designing a classification model done by you. Will you be proud of your design? Justify your answer.

Ans: A 100% training accuracy most of the time suggests overfitting. It is not necessarily a sign of a great model. One should check generalization error using cross validation scheme to ensure that the model generalizes well.

- c) What is the purpose of feature scaling in machine learning.

The main purpose of scaling in machine learning is to bring all feature values in the same numerical ranges. Widely used feature scaling methods are min-max normalization or z-score normalization. Following are the advantages of feature scaling:

1. Prevents large-valued features from dominating the learning process
2. Speeds up convergence for gradient-based methods
3. Handles outliers more effectively
4. Prevents overflow or underflow errors in the variables in the computation

- d) Define log-odds function and what is the range of log-odds in logistic regression?

Ans: Logistic regression predicts

$$p = \sigma(\mathbf{w}^T \mathbf{x} + b)$$

$$\text{where } \sigma(\mathbf{w}^T \mathbf{x} + b) = \frac{1}{(1 + e^{-(\mathbf{w}^T \mathbf{x} + b)})}$$

The *log-odds* is

$$\left(\frac{p}{1-p} \right) = \mathbf{w}^T \mathbf{x} + b$$

The main purpose of using *log-odds* in logistic regression is to span $(\mathbf{w}^T \mathbf{x} + b)$ as $(-\infty \text{ to } \infty)$.

- e) A dataset has 10 instances, where 6 belong to Spam Class and 4 belongs to Not Spam Class. Compute the entropy of the given dataset?

Ans: Entropy Calculation for the Given Dataset as follows:

$$H(D) = - \sum_{i=1}^c p_i \log_2(p_i)$$

where $H(D)$ is the entropy of the dataset; c is the number of classes; p_i is the probability of each class.

Step 1: Compute Class Probabilities

Spam Class: 6 instances

Not Spam Class: 4 instances

Total instances: 10

$$p(\text{Spam}) = \frac{6}{10} = 0.6 \text{ and } p(\text{Not Spam}) = \frac{4}{10} = 0.4$$

Step 2: Compute Entropy

$$H(D) = -[0.6 * \log_2(0.6) + 0.4\log_2(0.4)] = 0.97$$

So, the entropy of the given dataset is 0.97.

Note: As desired by the faculty, step marks should be awarded if at least some words in the answer of the student match with evaluation scheme. Otherwise award 0 mark.

2. You are given the following data from a simple linear regression model, where y_i represents the predicted values and \hat{y}_i represents the true values:

y_i	-114	-36.5	86	40
\hat{y}_i	-123	-36	122	50

Evaluate the performance of linear regression, calculates the residuals, MAE, MSE, RMSE, R-squared (R^2) value, and adjusted R-squared value. Based on the performance metric values provide a comment on whether the model is a good fit for the dataset.

Ans: The given data as follows:

Given Data:

y_i	-114	-36.5	86	40
\hat{y}_i	-123	-36	122	50

Step 1: Calculate Residuals

Residuals are the differences between the true values (y_i) and the predicted values (\hat{y}_i):

$$\text{Residual}_i = y_i - \hat{y}_i$$

y_i	\hat{y}_i	Residual ($y_i - \hat{y}_i$)
-114	-123	9
-36.5	-36	-0.5
86	122	-36
40	50	-10

Step 2: Calculate MAE (Mean Absolute Error)

MAE is the average of the absolute values of the residuals:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$\text{MAE} = \frac{|9| + |-0.5| + |-36| + |-10|}{4} = \frac{9 + 0.5 + 36 + 10}{4} = \frac{55.5}{4} = 13.875$$

Step 3: Calculate MSE (Mean Squared Error)

MSE is the average of the squared residuals:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\text{MSE} = \frac{9^2 + (-0.5)^2 + (-36)^2 + (-10)^2}{4} = \frac{81 + 0.25 + 1296 + 100}{4} = \frac{1477.25}{4} = 369.3125$$

Step 4: Calculate RMSE (Root Mean Squared Error)

RMSE is the square root of MSE:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{369.3125} \approx 19.22$$

Step 5: Calculate R-squared (R^2)

R-squared is the proportion of the variance in the dependent variable that is predictable from the independent variable(s):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

First, calculate the mean of the true values (\bar{y}):

$$\bar{y} = \frac{-114 + (-36.5) + 86 + 40}{4} = \frac{-24.5}{4} = -6.125$$

Now, calculate R-squared:

$$R^2 = 1 - \frac{1477.25}{23171.9375} \approx 1 - 0.0638 = 0.9362$$

Step 6: Calculate Adjusted R-squared

Adjusted R-squared adjusts the R-squared value based on the number of predictors and the sample size:

$$\text{Adjusted } R^2 = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$$

For simple linear regression, $k = 1$:

$$\text{Adjusted } R^2 = 1 - \left(\frac{(1 - 0.9362)(4 - 1)}{4 - 1 - 1} \right) = 1 - \left(\frac{0.0638 \times 3}{2} \right) = 1 - 0.0957 = 0.9043$$

The R-squared value of 0.9362 indicates that approximately 93.62% of the variance in the dependent variable is explained by the model, which is quite high. The adjusted R-squared value of 0.9043 also suggests a good fit, accounting for the number of predictors.

Comment: The linear regression model is a good fit for the dataset, as indicated by the high R-squared and adjusted R-squared values and relatively low error metrics (MAE, MSE, RMSE).

Note: For comment award 1 mark and rest as desired by the faculty, step marks should be awarded.

3. Consider the following dataset:

Income Level	Credit Score	Loan Amount	Default
Low	Bad	Small	Yes
Low	Bad	Large	Yes
Medium	Average	Medium	No
High	Good	Large	No
High	Average	Medium	No
Medium	Good	Small	No
Low	Average	Medium	Yes
High	Bad	Large	No
Low	Good	Medium	No

Using the Naive Bayes classifier, determine whether a new customer with the following attributes is likely to default on a loan: Income Level = Low, Credit Score = Average, Loan Amount = Medium.

Q13 Using Naive Bayes classifier, determine whether a new customer with the following attributes is likely to default on a loan.
sol Income Level = Low, credit score = Average, Loan = medium

step 1 compute prior probabilities

$$P(\text{Default} = \text{Yes}) = \frac{3}{9} = 0.333$$

$$P(\text{Default} = \text{No}) = \frac{6}{9} = 0.667$$

step 2 conditional probabilities for Default = Yes

$$P(\text{Income Level} = \text{Low} | \text{Default} = \text{Yes}) = \frac{3}{3} = 1$$

$$P(\text{Credit score} = \text{Avg} | \text{DEF} = \text{Yes}) = \frac{1}{3} = 0.333$$

$$P(\text{Loan amount} = \text{medium} | \text{Def} = \text{Yes}) = \frac{1}{3} = 0.333$$

Default = No

$$P(\text{Low} | \text{No}) = \frac{1}{6} = 0.1667$$

$$P(\text{Avg} | \text{No}) = \frac{2}{6} = 0.333$$

$$P(\text{medium} | \text{No}) = \frac{2}{6} = 0.333$$

steps posterior probabilities

$$P(\text{Yes} | X) = P(\text{Yes}) P(X | \text{Yes})$$

$$= 0.333 \times 1 \times 0.333 \times 0.333$$

$$\boxed{P(\text{Yes} | X) = 0.037}$$

→ Answer

Default = Yes

step 3 $P(\text{No} | X) = P(\text{No}) P(X | \text{No})$

$$= 0.667 \times 0.1667 \times 0.333 \times 0.333$$

$$\boxed{P(\text{No} | X) = 0.0185}$$

Yes > No

4. A retail company wants to classify new customers based on their annual income and spending behavior. The goal is to identify whether a customer is a Low Spender or a High Spender to tailor marketing strategies accordingly. The dataset below represents existing customers

Annual Income (in 1000\$)	Spending Score	Category
15	39	Low Spender
16	81	High Spender
17	6	Low Spender
18	77	High Spender
19	40	Low Spender

Given a new customer with Annual Income = \$17,000 and Spending Score = 50, Classify this new customer using KNN with k = 3. Use the Euclidean distance for calculations.

Ans:

Calculate Euclidean distances:

Distance to (15, 39): $\sqrt{(17-15)^2 + (50-39)^2} = \sqrt{4 + 121} = \sqrt{125} \approx 11.18$

Distance to (16, 81): $\sqrt{(17-16)^2 + (50-81)^2} = \sqrt{1 + 961} = \sqrt{962} \approx 31.00$

Distance to (17, 6): $\sqrt{(17-17)^2 + (50-6)^2} = \sqrt{0 + 1936} = \sqrt{1936} = 44.00$

Distance to (18, 77): $\sqrt{(17-18)^2 + (50-77)^2} = \sqrt{1 + 729} = \sqrt{730} \approx 27.02$

Distance to (19, 40): $\sqrt{(17-19)^2 + (50-40)^2} = \sqrt{4 + 100} = \sqrt{104} \approx 10.20$

Nearest Neighbors (k=3):

1. (19, 40) → Low Spender (10.20)

2. (15, 39) → Low Spender (11.18)

3. (18, 77) → High Spender (27.02)

Majority Category: 2 Low Spenders vs 1 High Spender

Hence, the new customer is classified as a Low Spender.

Note: As desired by the faculty, step marks should be awarded.

5. Find the distance from the point $[1 \ 1 \ 1 \ 1 \ 1]^T$ to the hyperplane

$$x_1 - x_2 + x_3 - x_4 + x_5 + 1 = 0 \quad [1 \text{ Mark}]$$

(a) Distance from $\mathbf{x}_0 = [1, 1, 1, 1, 1]^T$ to the plane

$$x_1 - x_2 + x_3 - x_4 + x_5 + 1 = 0.$$

Rewrite as $\mathbf{w} \cdot \mathbf{x} + b = 0$, with

$$\mathbf{w} = (1, -1, 1, -1, 1), \quad b = 1.$$

Compute

$$\mathbf{w} \cdot \mathbf{x}_0 + b = (1)(1) + (-1)(1) + (1)(1) + (-1)(1) + (1)(1) + 1 = 2,$$

and

$$\|\mathbf{w}\| = \sqrt{1^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2} = \sqrt{5}.$$

Hence the perpendicular distance is

$$\text{distance} = \frac{|\mathbf{w} \cdot \mathbf{x}_0 + b|}{\|\mathbf{w}\|} = \frac{2}{\sqrt{5}}.$$

Note: If a student has calculated up to norm of the weight vector, he/she should award 0.5 mark.

Explain the primal and dual formulation of the Support Vector Machine (SVM) optimization problem.

(b) Primal/Dual Formulation of SVM (Hard-margin) **Primal:** Given linearly separable data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with $y_i \in \{+1, -1\}$:

$$\min_{\mathbf{w}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \forall i.$$

Minimizing $\frac{1}{2} \|\mathbf{w}\|^2$ maximizes the margin $\frac{2}{\|\mathbf{w}\|}$.

Dual: Introducing Lagrange multipliers $\alpha_i \geq 0$, the dual problem is:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j),$$

subject to

$$\alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0.$$

Solving the dual yields the same optimum \mathbf{w} , with only *support vectors* having nonzero α_i .

[4 Marks]

Note: Here short derivation has given. However, complete derivation must be derived by the student to award 4 marks. Based on derivation, the students should be awarded marks out of 4.