NPTEL Online Certification Courses

Indian Institute of Technology Kharagpur

MACHINE LEARNING FOR SOIL AND CROP MANAGEMENT

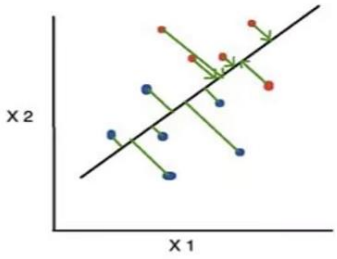
Assignment- Week 4

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 15 **Total mark: 15 X 1 = 15**

QUESTION 1:

1. The graph below represents:



a. Linear discriminant analysis

b. Linear regression analysis

c. Both a and b

d. None of the above

Correct Answer: a

Detailed Solution: the plot above shows the linear discriminant analysis.

QUESTION 2:

What is the primary goal of classification in machine learning?

a. To order the data

b. To identify the category of an observation

c. To maximize variance within data groups

d. To reduce dimensionality of data

Correct Answer: b

Detailed Solution: Classification assigns an observation to one of several predefined categories based on explanatory features.

QUESTION 3:

The formula below is the expression of:

$$d_{euc}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

a. Euclidean distance
b. Manhattan distance
c. Both a and b
d. None of the above

Correct Answer: a

Detailed Solution: The formula above represents Euclidean distance.

QUESTION 4:

What are the criteria used by LDA (Linear Discriminant Analysis) to create new axis?

a. Maximize the distance between means of the two classes and also maximize the variation within each class
b. Minimize the distance between means of the two classes and also minimize the variation within each class
c. Minimize the distance between means of the two classes and maximize the variation within each class
d. Maximize the distance between means of the two classes and minimize the variation within each class

Correct Answer: d

Detailed Solution: Criteria used by LDA to create new axis are 1. Maximize the distance between means of the two classes, 2. Minimize the variation within each class.

QUESTION 5:

_____ metrics represent the number of incorrect positive prediction out of total true negatives.

a. True negative
b. True positive
c. False positive
d. True positive

Correct Answer: c

Detailed Solution: False positive metrics represent the number of incorrect positive predictions out of total true negatives.

QUESTION 6:

Which of the following statements is correct for the ROC curve?

a. The ROC curve is a commonly used graph that summarizes the performance of a classifier over all possible thresholds.
b. It is generated by plotting the True Positive Rate (x-axis) against the False Positive Rate (y-axis) as we vary the threshold for assigning observations to a given class.
c. It is generated by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis) as we vary the threshold for assigning observations to a given class.
d. Both a and c

Correct Answer: d

Detailed Solution: ROC curve is a commonly used graph that summarizes the performance of a classifier over all possible thresholds. It is generated by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis) as we vary the threshold for assigning observations to a given class.

QUESTION 7:

_____ metrics represent the number of incorrect positive prediction out of total true negatives.

a. True negative
b. True positive
c. False positive
d. True positive

Correct Answer: c

Detailed Solution: False positive metrics represent the number of incorrect positive predictions out of total true negatives.

QUESTION 8:

What does the term "support vectors" refer to in SVM?

a. Data points farthest from the hyperplane
b. Data points closest to the hyperplane
c. All data points in the dataset
d. The coefficients of the hyperplane equation

Correct Answer: b

Detailed Solution: Support vectors are the data points nearest to the hyperplane, influencing its position and orientation.

QUESTION 9:

Which type of learning algorithm is K-Nearest Neighbors (KNN)?

a. Supervised learning
b. Unsupervised learning
c. Semi-supervised learning
d. Reinforcement learning

Correct Answer: a

Detailed Solution: K-Nearest Neighbors (KNN) is a supervised learning algorithm.

QUESTION 10:

Which of the following method is most commonly used to calculate the distance between test data and training data?

a. Hamming distance
b. Euclidean distance
c. Manhattan distance
d. Minkowski distance

Correct Answer: b

Detailed Solution: Euclidean distance method is most commonly used to calculate the distance between test data and training data.

QUESTION 11:

What does the term "hyperplane" refer to in SVM?

a. A line that minimizes the error in regression
b. A decision boundary that separates classes
c. A cluster centroid in k-means
d. A hierarchical level in clustering

Correct Answer: b

Detailed Solution: In SVM, the hyperplane is the decision boundary that maximally separates classes.

QUESTION 12:

In k-means, what happens if k is too large?

a. Clusters become too less
b. Computation time decreases
c. Data points are over-clustered
d. Clusters become more homogeneous

Correct Answer: c

Detailed Solution: A very large k may result in over-clustering, where meaningful groupings are split unnecessarily.

QUESTION 13:

What is the correct equation for odds ratio, when p is the probability that the event Y occurs, p(Y=1)?

a. Odds ratio = p/(1-p)
b. Odds ratio = p/(1+p)
c. Odds ratio = (1-p)/p
d. Odds ratio = (1+p)/p

Correct Answer: a

Detailed Solution: In logit model, $\ln[p/(1-p)] = \alpha + \beta X + e$, **$p/(1-p)$** is the odds ratio, and $\ln[p/(1-p)]$ is logit.

QUESTION 14:

Which of the following is not a classification metrics?

a. Recall
b. Precision
c. F1 score
d. RMSE

Correct Answer: d

Detailed Solution: The classification metrics are: accuracy, recall(sensitivity or true positive rate), precision, false positive rate, true negative rate (specificity), F1 score, Matthews correlation coefficient (MCC), Cohen's kappa, etc.

QUESTION 15:

Which of the following statements is/are true for a good clustering method to produce high-quality clusters?

a. The intra-cluster similarity is high
b. The inter-class similarity is low
c. Able to discover some or all of the hidden patterns
d. All of the above

Correct Answer: d

Detailed Solution: A good clustering method will produce high quality clusters in which the intra-cluster similarity is high, the inter-class similarity is low, and able to discover some or all of the hidden patterns.