NPTEL Online Certification Courses

Indian Institute of Technology Kharagpur

MACHINE LEARNING FOR SOIL AND CROP MANAGEMENT

Assignment- Week 2

TYPE OF QUESTION: MCQ/MSQ

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

QUESTION 1:

What does each row in a data matrix represent?

a. A variable

b. A sample

c. A feature

d. A regression line

Correct Answer: b

Detailed Solution: Each row in a data matrix corresponds to a sample, while each column corresponds to a feature or variable.

QUESTION 2:

For any Simple Linear Regression (SLR) model of equation $Y_i = B_1X_i + B_0$, estimate of the Intercept is denoted by

a. Y_i
b. B_1
c. B_0
d. X_i

Correct Answer: c

Detailed Solution: Y_i = dependent variable, B_1 = Slope/Coefficient, B_0 = Constant/Variable, X_i = Independent variable

QUESTION 3:

What is multivariate data?

a. A single column of variables
b. A single row of variables
c. A rectangular matrix with rows and columns of numerical values
d. None of the above

Correct Answer: c

Detailed Solution: Multivariate data consists of a rectangular matrix (spreadsheet) with rows as samples and columns as features (variables). Each cell contains a numerical value.

QUESTION 4:

Why is centering and scaling data important?

a. To introduce skewness
b. To reduce mean to zero and variance to one
c. To increase multi-collinearity
d. To remove outliers

Correct Answer: b

Detailed Solution: Centering and scaling ensure data has a mean of zero and a standard deviation of one.

QUESTION 5:

What is an outlier most likely to affect in a regression analysis?

a. Slope only
b. Correlation coefficient
c. Mean value of residuals
d. None of the above

Correct Answer: b

Detailed Solution: Outliers can significantly affect the correlation coefficient and regression line due to their extreme values.

QUESTION 6:

Which plot is used to visualize the relationship between two variables?

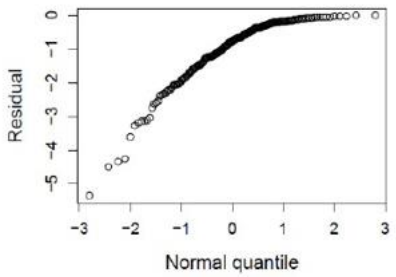
a. Histogram
b. Scatterplot
c. Boxplot
d. Line plot

Correct Answer: b

Detailed Solution: Scatter plots are used to visualize the relationship and correlation between two variables.

QUESTION 7:

How can you describe the plot below?



a. Left skew
b. Right skew
c. Normal
d. Heavy tails

Correct Answer: a

Detailed Solution: The above plot represents the left-skewed distribution, with most of the data falling to the left.

QUESTION 8:

Which of the following is a common tool to deal with nonlinearity, non-normality and unequal error variance problem in linear regression analysis?

a. Multivariate classification
b. Clustering
c. Transformation
d. Multivariate Calibration

Correct Answer: c

Detailed Solution: Transformation is a common tool to deal with nonlinearity, non-normality and unequal error variance problems in linear regression analysis.

QUESTION 9:

What is the purpose of early stopping in machine learning?

a. To introduce multicollinearity

b. To prevent overfitting by halting training early

c. To standardize data

d. To increase bias

Correct answer: b

Detailed answer: Early stopping halts training before the model learns noise, preventing overfitting.

QUESTION 10:

What is a key advantage of bagging?

a. Increases variance within noisy dataset

b. Reduces overfitting in high-dimensional data

c. Increases model complexity

d. Enhances the effect of outliers

Correct Answer: b

Detailed Solution: The advantages of bagging are: reduction in over-fitting of the model, handles higher dimensionality data very well, and maintains accuracy for missing data.

QUESTION 11:

Which of the following is the feature of underfitting?

a. High bias, high variance
b. High bias, low variance
c. Low bias, high variance
d. Low bias, low variance

Correct Answer: b

Detailed Solution: Underfit: high bias, low variance.

QUESTION 12:

Which of the following way can help in avoiding overfitting?

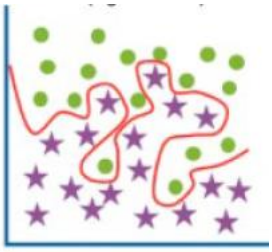
a. Early stopping of the model
b. Including more data in the training model
c. Feature selection
d. All of the above

Correct Answer: d

Detailed Solution: The overfitting can be avoided by various ways. For example, by early stopping of the model, by including more data in the training model, data augmentation, feature selection, regularization, and ensemble method.

QUESTION 13:

How you can describe the model trend in the below plot?



a. Underfitting, high bias
b. Overfitting, high variance
c. Optimum
d. None of the above

Correct Answer: b

Detailed Solution: The plot above represents overfitting with high variance.

QUESTION 14:

Which method is least likely to be affected by multicollinearity?

a. Polynomial regression
b. Simple linear regression
c. Random forest
d. Support Vector Machines

Correct Answer: c

Detailed Solution: Tree-based algorithms like Random Forest are not significantly impacted by multicollinearity as they split on features with higher information gain.

QUESTION 15:

What is the significance of the intercept in a simple linear regression model?

a. It represents the maximum value of the dependent variable.
b. It is the predicted value of the dependent variable when the independent variable is 0.
c. It measures the strength of the relationship between variables.
d. It is the difference between observed and predicted values.

Correct Answer: b

Detailed Solution: It is the predicted value of the dependent variable when the independent variable is 0. The intercept shows where the regression line crosses the y-axis.