Jane Chinwuko
Project: Batch DS2307
PFA Worksheet Set 1

Machine Learning.

Q1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B

Answer is A

The least Square Error is a curve whose sum of the squared residuals (or deviations or errors) from the given data points is at a minimum.

Q2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers
B) linear regression is not sensitive to outliers
C) Can't say
D) none of these

Answer is A

An outlier is a value in a dataset that deviates greatly from the normal distribution and is considered abnormal. They have the potential to significantly skew any regression model.

Q3. A line falls from left to right if a slope is _____?

A) Positive
B) Negative
C) Zero
D) Undefined

Answer is B

If a line's slope moves downward as we move from left to right, that slope is said to have a negative slope.

Q4. Which of the following will have symmetric relation between dependent variable and independent variable?

A) Regression
B) Correlation

C) Both of them
D) None of these

Answer is B

Correlation is a statistical approach that shows how one variable moves or changes in relation to another variable.

Q5. Which of the following is the reason for over fitting condition?

A) High bias and high variance
B) Low bias and low variance
C) Low bias and high variance
D) none of these

Answer is C

A model that overfits the target is one with low bias and high variance.

Q6. If output involves label then that model is called as:

A) Descriptive model
B) Predictive modal
C) Reinforcement learning
D) All of the above

Answer is B

Based on the input data, a predictive model for classification would predict a distinct class label as output.

Q7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularisation

Answer is D

Regularisation is a technique used in machine learning to reduce errors by properly fitting the function on a training set and preventing overfitting. Lasso and Ridge regression are regularisation techniques.

Q8. To overcome with imbalance dataset which technique can be used?

A) Cross validation

B) Regularisation
C) Kernel
D) SMOTE

Answer is D

One of the oversampling techniques most frequently used to address the imbalance dataset issue is SMOTE (synthetic minority oversampling technique).

Q9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

A) TPR and FPR
B) Sensitivity and precision
C) Sensitivity and Specificity
D) Recall and precision

Answer is A

The TPR(True Positive Rate) indicates the proportion of correct positive results in all positive samples made available for the test.

When comparing all available negative samples during the test, FPR(False Positive Rate) determines how frequently false positive results occur.

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True

B) False

Answer is B

The model's performance in separating the positive and negative classes improves with increasing AUC.

Q11. Pick the feature extraction from below:

A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Answer is B

Principal component analysis (PCA), a technique for reducing the dimensionality of large data sets, transforms a large collection of variables into a smaller set that retains most of the information in the larger set.

Q12. In Q12, more than one options are correct, choose all the correct options:

Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.
D) It does not make use of dependent variable.

Answer is A and B

Q13 and Q15 are subjective answer type questions, Answer them briefly.

Q13. Explain the term regularisation?

Regularisation is a technique used in machine learning to reduce errors by properly fitting the function on a training set and preventing overfitting. This process improves the performance of the model.

The regularisation process can be carried out in different ways, but it usually involves the addition of penalties or constraints on the optimisation process.

Q14. Which particular algorithms are used for regularisation?

There are three commonly used regularisation algorithms which are:

1. **Lasso Regularization – L1 Regularization**
   When using Lasso regularisation, a penalty is added to the absolute value of the magnitude of the coefficients.

2. **Ridge Regularization – L2 Regularization**
   For Ridge regularisation, a penalty is added to the square of the magnitude of the coefficients.

3. **Elastic-Net Regularization – L1 and L2 Regularization**
   Elastic Net regularisation combines both Lasso and Ridge regularisation.

Q15. Explain the term error present in linear regression equation?

The error term represents the difference between the true values and observed values in a population data. This error term aids in calculating the R-squared value, which indicates how effective the model is on the whole.