

# Machine Learning

Section – A (Only one correct option)

Ans -1 (A) Least Square Error

Ans -2 (A) Linear regression is sensitive to outliers

Ans -3 (B) Negative

Ans -4 (B) Correlation

Ans -5 (C) Low bias and high variance

Ans -6 (B) Predictive model

Ans -7 (D) Regularization

Ans -8 (D) SMOTE

Ans -9 (A) TPR and FPR

Ans -10 (B) False

Ans -11 (A) Construction bag of words from an email

Section – B (More than one correct option)

Ans -12

(B) It becomes slow when number of features is very large

(D) It does not make use of dependent variable

Section – C (Subjective)

Ans -13 When a machine/deep learning model is built, it has some biasness and variance towards data while training, which may affect the results while using the same model on test data. This situation leads to either underfitting or overfitting of the model. Also, it may behave differently to entirely new set of data. In order to solve this issue, generalisation of the model is required which could be achieved by the method of regularization. The process of regularization helps to minimise the loss function which effectively leads to minimising the problem of underfitting and overfitting.

Ans -14 Regularization of machine learning model could be done in three ways and contains different methods under each of them, few of them are:-

Method -1: Modifying loss function:

1. L1 Norm (Lasso Regularization)
2. L2 Norm (Ridge Regularization)

### 3. Entropy Regularization

Method -2: Modifying sampling method:

1. Data Augmentation
2. K-fold Cross-Validation

Method -3: Modifying training algorithm:

1. Dropout
2. Injecting noise

Ans -15 The error term present in linear regression equation is symbolised as  $\epsilon$ . It is the difference between the actual data value and the predicted data value by the model. However, it is not the actual error of the model, in order to find out the error of the model, we have to square the distances between the actual and predicted values and then take the mean of each squared distances.