MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. 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 : A) Least square Error

Line of best fit refers to a line through a scatter plot of data points that best expresses the relationship between those points.

The most precise method involves least square error methods.

The least Sum of Squares of Errors is used as the cost function for Linear Regression. For all possible lines, calculate the sum of squares of errors. The line which has the least sum of squares of errors is the best fit line

2. 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

Ans : A) Linear regression is sensitive to outliers

The slope of the regression line will change due to outliers in most cases. So , Linear Regression is sensitive to outliers.

3. A line falls from left to right if a slope is _____?
A) Positive
B) Negative
C) Zero
D) Undefined

Ans: B) Negative

The slope of a line measures the steepness of the line.

Slope = Rise/ Run.

Rise means how many units you move up or down from point to point. On the graph that would be a change in the y values.

Run means how far left or right you move from point to point. On the graph, that would mean a change of x values.
Slope = -2/1 = -ve
Note that when a line has a negative slope it falls left to right.

4. 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 : C)

5. 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 : C) Low bias and high variance

1. Low-Bias, Low-Variance:
   The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.
2. Low-Bias, High-Variance: With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an overfitting
3. High-Bias, Low-Variance: With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to underfitting problems in the model.
4. High-Bias, High-Variance:
   With high bias and high variance, predictions are inconsistent and also inaccurate on average.

6. 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 : B) Predictive Model

The descriptive analysis uses mainly unsupervised learning approaches for summarizing, classifying, extracting rules to **answer what happens was happened in the past.**

Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data.A predictive model based on supervised learning algorithms. **A supervised machine learning model will learn to identify patterns and relationships within a labelled training dataset.**

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, **take actions and learn through trial and error.**

7. Lasso and Ridge regression techniques belong to _____?

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

Answer : D) Regularisation

The commonly used regularization techniques are :

Lasso Regularization – L1 Regularization
Ridge Regularization – L2 Regularization

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

A) Cross validation
B) Regularisation
C) Kernel
D) SMOTE

Answer : D) SMOTE

Imbalanced data refers to those types of datasets where the target class has an uneven distribution of observations, i.e one class label has a very high number of observations and the other has a very low number of observations.

More such example of imbalanced data is –

· Disease diagnosis
· Fraud detection
· Natural disaster

There are many techniques available to overcome imbalanced data , one of them is SMOTE

SMOTE

Synthetic Minority Oversampling Technique or SMOTE is another technique to oversample the minority class. Simply adding duplicate records of minority class often don't add any new information to the model. In SMOTE new instances are synthesized from the existing data. If we explain it in simple words, SMOTE looks into minority class instances and use k nearest neighbor to select a random nearest neighbor, and a synthetic instance is created randomly in feature space.

9. 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 : A)

AUC – Area Under Curve
ROC – Receiver Operating characteristics

The Receiver Operator Characteristic (ROC) curve is an evaluation metric for binary classification problems. It is a probability curve that plots the TPR against FPR at various threshold values
The Area Under the Curve (AUC) is the measure of the ability of a binary classifier to distinguish between classes and is used as a summary of the ROC curve.
The value which is obtained is equal to the area under the receiver operating characteristics curve. Its value remains between 0 to 1.

TPR – True Positive Rate
FPR – False Positive Rate

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

A) True
B) False

Ans : B) False

The higher the AUC, the better the model's performance at distinguishing between the positive and negative classes.

11. 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

Ans : B) Apply PCA to project high dimensional data

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

12. 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.

Ans : A), B) and C)

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

13. Explain the term regularization?

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it. It basically reduces or regularizes the coefficient of features towards zero.

Overfitting is a phenomenon that occurs when a Machine Learning model is constrained to the training set and not able to perform well on unseen data. That is when our model learns the noise in the training data as well. This is the case when our model memorizes the training data instead of learning the patterns in it.

Underfitting on the other hand is the case when our model is not able to learn even the basic patterns available in the dataset. In the case of the underfitting model is unable to perform well even on the training data hence we cannot expect it to perform well on the validation data. This is the case when we are supposed to increase the complexity of the model or add more features to the feature set.

Working of Regularisation :
Regularization works by adding a penalty or complexity term to the complex model. Let's consider the simple linear regression equation:
$y= \beta0+\beta1x1+\beta2x2+\beta3x3+\cdots+\beta nxn +b$
In the above equation, Y represents the value to be predicted
X1, X2, …Xn are the features for Y.
$\beta0,\beta1,\ldots..\beta n$ are the weights or magnitude attached to the features, respectively. Here represents the bias of the model, and b represents the intercept.
Linear regression models try to optimize the β0 and b to minimize the cost function. The equation for the cost function for the linear model is given below:

$$\sum_{i=1}^{M}\left(y_i - y'_i\right)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * Xij\right)^2$$

Now, we will add a loss function and optimize parameter to make the model that can predict the accurate value of Y. The loss function for the linear regression is called as **RSS or Residual sum of squares.**

The commonly used regularization techniques are :

- **Lasso Regularization – L1 Regularization**
- **Ridge Regularization – L2 Regularization**

14. Which particular algorithms are used for regularization?

The commonly used regularization techniques are :

Lasso Regularization – L1 Regularization

Ridge Regularization – L2 Regularization

Ridge Regularization :

- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.
- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called **Ridge Regression penalty**. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.
- The equation for the cost function in ridge regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}\beta_j{}^2$$

- In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model.
- As we can see from the above equation, if the values of **λ tend to zero, the equation becomes the cost function of the linear regression model.** Hence, for the minimum value of λ, the model will resemble the linear regression model.
- A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
- It helps to solve the problems if we have more parameters than samples.

Lasso Regression:

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for **Least Absolute and Selection Operator.**
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as **L1 regularization.** The equation for the cost function of Lasso regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|^{\square}$$

- Some of the features in this technique are completely neglected for model evaluation.
- Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

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

Linear Regression is a supervised statistical technique where we try to estimate the dependent variable with a given set of independent variables.

We assume the relationship to be linear and our dependent variable must be continuous in nature.

The red line is the fitted line of regression and the points denote the actual observations.The vertical distance between the points and the fitted line (line of best fit) are called errors.

The main idea is to fit this line of regression by minimizing the sum of squares of these errors. This is also known as principle of least squares.

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon$$

Here,
'y' is the dependent variable to be estimated
X are the independent variables and
ε is the error term.

The error term includes everything that separates your model from actual reality.

This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.
It **is the deviation between observed points and the fitted line,** represents the margin of error within a statistical model.

It refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.

Zero conditional mean is there which says that there are both negative and positive errors which cancel out on an average. This helps us to estimate dependent variable precisely.