

1.The correct answer is D) Both A and B.

In Linear Regression, we use two methods to find the best fit line for data:

Least Square Error (LSE): This method involves minimizing the sum of the squared differences between the observed and predicted values. The goal is to find the line that minimizes the sum of the squared errors.

Maximum Likelihood (ML): This method involves finding the parameters that maximize the likelihood of observing the data given the model. In the case of Linear Regression, the likelihood function is a Gaussian distribution, and the ML estimate is equivalent to the LSE estimate.

2.The correct answer is A) Linear regression is sensitive to outliers.

Outliers are data points that are significantly different from the other data points in the dataset. In Linear Regression, outliers can have a significant impact on the model's performance and accuracy.

3.The correct answer is B) Negative.

The slope of a line represents how steep it is and in which direction it slopes. A negative slope indicates that the line slopes downward from left to right, meaning that as the x-coordinate increases, the y-coordinate decreases.

4.The correct answer is B) Correlation.

Correlation measures the strength and direction of the linear relationship between two continuous variables, typically denoted as X (independent variable) and Y (dependent variable). Correlation is symmetric, meaning that the correlation between X and Y is the same as the correlation between Y and X. In other words, the correlation coefficient (e.g., Pearson's r) remains the same regardless of which variable is considered the independent variable and which is considered the dependent variable.

5.The correct answer is C) Low bias and high variance.

Overfitting occurs when a model is too complex and learns the noise in the training data, rather than the underlying patterns. This results in a model that performs well on the training data but poorly on new, unseen data.

6.The correct answer is B) Predictive model.

A predictive model is a type of machine learning model that makes predictions about a target or response variable based on input features or predictor variables. The output of a predictive model involves a label or a predicted value, which is the expected outcome or response.

7.The correct answer is D) Regularization.

Regularization techniques add a penalty term to the loss function to discourage large weights or coefficients. This helps to prevent the model from fitting the noise in the data too closely.

8.The correct answer is D) SMOTE.

SMOTE (Synthetic Minority Over-sampling Technique) is a technique used to overcome the problem of imbalance datasets, where one class has a significantly larger number of instances than the other class(es). SMOTE works by creating synthetic samples of the minority class, thereby increasing its size and balancing the dataset.

9.The correct answer is A) TPR and FPR.

The AUC-ROC (Area Under the Receiver Operating Characteristic) curve is a graphical representation of the performance of a binary classification model. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at different thresholds.

True Positive Rate (TPR) is the ratio of true positives to the sum of true positives and false negatives.

False Positive Rate (FPR) is the ratio of false positives to the sum of false positives and true negatives.

10.The correct answer is B) False.

In an AUC-ROC (Area Under the Receiver Operating Characteristic) curve, a better model is indicated by a higher area under the curve, not a lower one. The area under the curve represents the model's ability to distinguish between positive and negative classes.

11.The correct answer is A) Construction bag of words from a email and C) Removing stop words.

Feature extraction is the process of extracting relevant information or features from raw data, such as text, images, or audio, to reduce the dimensionality and improve the quality of the data for modeling.

A) Construction bag of words from an email: This is a feature extraction technique in natural language processing (NLP) where a document (in this case, an email) is represented as a bag, or a set, of its word frequencies. This is a common technique used in text classification, sentiment analysis, and topic modeling.

12.The correct answer is A) We don't have to choose the learning rate.

The Normal Equation is a method for computing the coefficients of a Linear Regression model. It is a closed-form solution that finds the optimal values of the coefficients by minimizing the sum of the squared errors.

13.Regularization is a technique used in machine learning and statistics to prevent overfitting, which occurs when a model is too complex and performs well on the training data but poorly on new, unseen data.

In essence, regularization adds a penalty term to the loss function of the model, which discourages the model from having large weights or complex relationships between variables. This helps to:

Reduce overfitting: By adding a penalty term, the model is forced to find a simpler solution that generalizes better to new data.

Improve model interpretability: Regularization can help to identify the most important features and reduce the impact of noisy or irrelevant variables.

Enhance model robustness: Regularization can make the model more robust to outliers, noisy data, and changes in the data distribution.

14.Ridge Regression:

2.Lasso Regression:

3.Elastic Net Regression:

4.Dropout:

5.Early Stopping:

6.Gradient Penalty:

7.Weight Decay:

8.Batch Normalization:

9.L1 and L2 Regularization for Neural Networks:

10.Tikhonov Regularization:

15.In linear regression, the error term, also known as the residual or disturbance term, is a crucial component of the equation. It represents the difference between the observed value of the dependent variable (y) and the predicted value of the dependent variable based on the independent variables (x).

The linear regression equation is typically written as:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\mu}$$

Where:

y is the dependent variable (outcome or response variable)

x is the independent variable (predictor or feature variable)

$\hat{\beta}_0$ is the intercept or constant term

$\hat{\beta}_1$ is the slope coefficient

$\hat{\mu}$ is the error term (residual or disturbance term)