MACHINE LEARNING

Ans-1] D) Both A and B

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 modal

Ans-7] D) Regularization

Ans-8] D) SMOTE

Ans-9] A) TPR and FPR

Ans-10] B) False

Ans-11] B) Apply PCA to project high dimensional data

Ans-12] A) We don't have to choose the learning rate.

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

Ans-13

Regularization is a technique used in machine learning and statistical modeling to prevent overfitting and improve the generalization ability of a model. Overfitting occurs when a model fits the training data too closely, capturing noise and random fluctuations in the data, which leads to poor performance on unseen data.
Regularization introduces additional constraints or penalties to the model's objective function during the training process. These constraints encourage the model to have simpler and smoother representations, reducing the complexity and variance of the learned model. Regularization helps strike a balance between fitting the training data well and avoiding overfitting.
The most common types of regularization are:
L1 Regularization (Lasso): This technique adds the absolute values of the model's coefficients as a penalty term to the objective function. It encourages sparsity by driving some of the coefficients to exactly zero, effectively performing feature selection. L1 regularization is useful when we have a high-dimensional dataset and want to identify the most important features.
L2 Regularization (Ridge): L2 regularization adds the squared magnitudes of the model's coefficients as a penalty term to the objective function. It encourages small but non-zero coefficients for all features, effectively shrinking them towards zero. L2 regularization is effective for reducing the impact of less important features and improving the stability of the model.
Elastic Net Regularization: Elastic Net combines L1 and L2 regularization by adding both penalties to the objective function. It provides a balance between feature selection (L1 regularization) and coefficient shrinkage (L2 regularization). Elastic Net is useful when dealing with highly correlated features.
Regularization parameters, such as the regularization strength (lambda or alpha), control the amount of penalty applied to the model's coefficients. The optimal value of the regularization parameter is typically determined through techniques like cross-validation.
By using regularization, models can generalize better to unseen data by reducing overfitting and making them less sensitive to noise and irrelevant features in the training data.

Ans-14

Several machine learning algorithms can utilize regularization techniques to improve their performance. Some of the commonly used algorithms that incorporate regularization are:

1) Ridge Regression: Ridge regression adds an L2 regularization term to the linear regression cost function. It helps to prevent overfitting and reduce the impact of individual features by shrinking the coefficients towards zero.

2) Lasso Regression: Lasso regression incorporates an L1 regularization term into the linear regression cost function. It promotes sparsity by encouraging some coefficients to be exactly zero, effectively performing feature selection.

3) lastic Net: Elastic Net combines both L1 and L2 regularization. It adds a combination of L1 and L2 penalty terms to the linear regression cost function, allowing for both feature selection and coefficient shrinkage.

4) Logistic Regression: Logistic regression can also benefit from regularization techniques. Regularized logistic regression applies L1 or L2 regularization to the logistic regression cost function, helping to prevent overfitting and improve generalization.

5) Support Vector Machines (SVM): SVM algorithms can be regularized using techniques such as L1 or L2 regularization. Regularization in SVM helps control the balance between maximizing the margin and minimizing the training error.

6) Neural Networks: Regularization techniques like L1 and L2 regularization can be applied to the weights of neural networks to prevent overfitting and improve generalization. This is often done through techniques known as weight decay or weight regularization.

Ans-15

In the context of linear regression, the term "error" refers to the difference between the predicted values of the dependent variable (output) and the actual observed values. It represents the discrepancy or deviation between the predicted response of the linear regression model and the true response in the dataset.

In linear regression, the goal is to find the best-fitting line or hyperplane that minimizes the overall error between the predicted values and the actual values. This error is also known as the residual or the difference between the observed value and the predicted value for each data point.

PYTHON WORKSHEET

Ans-1] C) %

Ans-2] B) 0

Ans-3] C) 24

Ans-4] A) 2

Ans-5] D) 6

Ans-6] C) the finally block will be executed no matter if the try block .

raises an error or not

Ans-7] A) It is used to raise an exception.

Ans-8] C) in defining a generator

Ans-9] A) _abc & C) abc2

Ans-10] A) yield & B) raise

STATISTICS WORKSHEET-1

Ans-1] a) True

Ans-2] Central Limit Theorem

Ans-3] b) Modeling bounded count data

Ans-4] d) All of the mentioned

Ans-5] c) Poisson

Ans-6] b) False

Ans-7] b) Hypothesis

Ans-8]a)0

Ans-9] c) Outliers cannot conform to the regression relationship

Ans-10

This statistical distribution with symmetrical bell-shaped curve, defined by mean and standard deviation parameters, is what the term “Normal Distribution” signifies. It is also known as the Gaussian distribution or the bell curve.

A normal distribution is defined by its probability density function (PDF), which describes the relative likelihood of observing different values. The PDF of a normal distribution is symmetric around its mean, which is the central value of the distribution. The standard deviation determines the spread or dispersion of the distribution.

The key features of a normal distribution are:

1) Symmetry: The distribution is symmetric around its mean, with equal probabilities of observing values to the left and right of the mean.

2)

Bell-shaped: The distribution forms a smooth, bell-shaped curve, with the majority of the data concentrated around the mean.

3)

Although the force that drives many real-world variables to cluster around the mean arises from myriad individually random and haphazard elements, the aggregate effect of these disparate vagaries coalesces into a pattern that we recognize as the familiar bell curve, by virtue of a mathematical principle known as the Central Limit Theorem whereby a multitude of individually chaotic and arbitrary factors , when combined in large number, will tend spontaneously and inexorably to organize themselves around a stable and balanced norm.

4

) Parameters: The normal distribution is fully characterized by its mean (μ) and standard deviation (σ), which determine the location and shape of the distribution, respectively.

The normal distribution is widely used in statistics and probability theory due to its mathematical tractability and its relevance to various natural and social phenomena. It serves as a foundation for many statistical techniques, such as hypothesis testing, confidence intervals, and regression analysis.

Ans-11

Taking care of missing information is a significant stage in information examination, and there are a few strategies accessible for crediting missing qualities. The decision of attribution strategy relies upon the idea of the information and the fundamental suppositions. The following are a couple of normally utilized attribution procedures:

- Mean/middle ascription: In this strategy, missing qualities are supplanted with the mean or middle of the accessible qualities for that variable. This strategy is easy to execute however can underrate the changeability and misshape the conveyance of the variable.
- Hot deck ascription: This strategy includes haphazardly choosing a worth from a comparative benefactor unit (one more perception with comparative qualities) and doling out it to the missing worth. This strategy saves the connections between factors yet may present some irregularity.
- Various attribution: Different ascription includes making numerous credited datasets, where missing qualities are attributed on various occasions in view of a measurable model. This strategy represents the vulnerability related with attributions and gives more precise evaluations contrasted with single ascription techniques.
- Relapse ascription: This approach includes utilizing relapse models to foresee missing qualities in light of different factors. A relapse model is fitted utilizing the accessible information, and the anticipated qualities are utilized to credit the missing qualities. This technique catches the connections between factors yet expects that the relationship is direct.
- Model-based ascription: Model-based attribution utilizes progressed procedures, for example, most extreme probability assessment or Bayesian techniques to credit missing qualities in light of a fitted factual model. This approach considers the distributional suspicions of the factors and can deliver precise attributions.

While picking an attribution strategy, it is critical to think about the suppositions of the procedure, the idea of the missing information, and the particular necessities of the investigation. It is additionally fitting to survey the effect of missing information and lead awareness examinations to assess the power of the outcomes to various attribution systems.

Ans-12

A/B testing, otherwise called split testing, is a measurable strategy used to look at least two varieties of a website page, ad, or different components to figure out which one performs better as far as accomplishing a particular goal. It is broadly utilized in advertising, client experience examination, and site improvement.

In A/B testing, at least two forms of a site page or component (alluded to as the control and treatment gatherings) are displayed to various gatherings of clients haphazardly. The clients are then noticed and their reactions are estimated, for example, navigate rates, transformation rates, or client commitment measurements. The object is to assess which variant prompts better execution in light of the picked measurement.

A/B testing considers information driven independent direction by giving experiences into which varieties lead to improved results. It advances plans, showcasing efforts, client encounters, and different components by iteratively refining and further developing in view of certifiable client conduct.

Ans-13

Mean attribution of missing information is a basic and direct way to deal with handle missing qualities by supplanting them with the mean of the accessible qualities for that variable. While mean ascription is generally utilized and simple to execute, it has limits and potential disadvantages that should be thought of.

Mean attribution can in any case be utilized in specific circumstances, for example, when the missingness is negligible or the factors are not exceptionally connected. Be that as it may, it is by and large prescribed to consider more modern attribution procedures, for example, different attribution or model-based ascription, which can give more precise and legitimate outcomes by representing the vulnerability and connections in the information.

Ans-14

The basic form of a linear regression model with one independent variable is expressed as:

$Y = \beta_0 + \beta_1 X + \varepsilon$

Where:

- Y is the dependent variable (also known as the response variable or the outcome variable).
- X is the independent variable (also known as the predictor variable or the explanatory variable).
- β_0 is the intercept (the value of Y when X is 0).
- β_1 is the slope (the change in Y for a one-unit change in X).
- ε is the error term, representing the random variation or unexplained variability in the relationship between X and Y.

The goal of linear regression is to estimate the values of β_0 and β_1 that minimize the sum of the squared differences between the observed Y values and the predicted values based on the linear equation. This estimation is typically done using the method of least squares.

Ans-15

Measurements is a wide field that includes different branches or sub-trains, each zeroing in on various parts of information examination, demonstrating, and deduction. A portion of the principal parts of insights include:

1) Clear Insights: Unmistakable insights includes summing up and portraying information utilizing measures like mean, middle, standard deviation, and graphical portrayals like histograms, box plots, and disperse plots. It centers around coordinating and introducing information to acquire bits of knowledge and comprehend the essential qualities of a dataset.
2) Inferential Measurements: Inferential insights includes reaching determinations and making inductions about populaces in light of test information. It utilizes factual procedures to appraise populace boundaries, test speculations, and evaluate

vulnerability. Strategies, for example, speculation testing, certainty stretches, and relapse examination are ordinarily utilized in inferential measurements.

3) Likelihood Hypothesis: Likelihood hypothesis is the underpinning of measurements and manages the evaluation and investigation of vulnerability. It gives a structure to demonstrating and breaking down irregular occasions and cycles. Likelihood hypothesis is fundamental for figuring out ideas like arbitrary factors, likelihood appropriations, and testing.
4) Biostatistics: Biostatistics is the use of measurable strategies to natural, clinical, and general wellbeing research. It includes planning tests, breaking down clinical preliminaries, evaluating dangers and results, and deciphering wellbeing related information. Biostatistics assumes a vital part in the study of disease transmission, hereditary qualities, clinical examination, and medical services strategy.
5) Econometrics: Econometrics applies factual strategies to financial information to study and investigate monetary connections and peculiarities. It includes demonstrating and breaking down financial factors, assessing monetary boundaries, and testing monetary hypotheses utilizing factual methods.

These are only a couple of instances of the parts of measurements, and there are other specific regions like social insights, modern insights, monetary insights, and that's just the beginning. The field of measurements is assorted and interdisciplinary, with applications in different logical, social, and business areas.