Answer File - Fliprobo || Machine Learning Assignment-39 || Task 4

Batch - DS2406
Name - Sakshi Jha

Ans-1. D. Both A and B

In linear Regression, we employ both Least Squares Error (LSE) and Maximum Likelihood Estimation (MLE) methods to determine the best-fit line for the data:

- **Least Squares Error:** This method minimizes the sum of the squared differences between the observed dependent variable (Y) and the predicted values from the linear equation. It aims to find the line that provides the best fit to the data by minimizing the residuals.
- **Maximum Likelihood Estimation:** MLE involves selecting parameters that maximize the likelihood of observing the data under a given statistical model. In the context of linear regression, MLE helps estimate the coefficients (slope and intercept) of the linear equation that best explains the observed data.

Both of these methods are fundamental in Linear Regression for finding the optimal parameters that define the best-fit line to the data set. Therefore, the correct answer to how we find the best-fit line in Linear Regression is that we utilize both Least Squares Error and Maximum Likelihood.

Ans-2. a. Linear regression is sensitive to outliers
Outliers can have a disproportionate impact on the estimated coefficients (slope and intercept) of the regression line. They can pull the line towards them, affecting its slope and potentially biasing the model's predictions.

Ans-3. b. Negative
A negative slope means that as you move from left to right along the line, the y-values (vertical axis) decrease. In other words, the line "falls" from left to right

The Poisson distribution is used for modeling count data that can take any non-negative integer value and is typically unbounded, meaning there is no upper limit on the count. Bounded count data, where the counts are limited to a certain range, is not appropriately modeled by the Poisson distribution.

Ans-4. B. Correlation.

Correlation is correct because it explicitly measures the symmetry in the relationship between variables, whereas regression (option A) is a modeling technique that estimates how the dependent variable changes as the independent variables change, and this relationship may not be symmetric in all cases.

Ans-5 b. Low bias and high variance

Bias refers to the error introduced by approximating a real-world problem with a simpler model. High bias means the model is too simplistic and fails to capture the underlying patterns in the data. While Variance refers to the model's sensitivity to small fluctuations in the training data. High variance means the model is overly complex and captures noise and random fluctuations in the training data.
Overfitting occurs when a model learns not only the underlying patterns in the data but also the noise and random fluctuations specific to the training set. This results in a model that performs well on the training data but fails to generalize to new, unseen data.

Ans-6. b. Predictive Model
A predictive model is a type of model that is trained to predict an output variable (often referred to as a label or target variable) based on input variables (features or predictors). The goal of a predictive model is to make accurate predictions about future or unseen data based on patterns learned from historical data.

Ans-7. D. Regularization
Lasso (Least Absolute Shrinkage and Selection Operator) and Ridge regression are both regularization techniques used in linear regression to prevent overfitting by adding a penalty term to the model's cost function. Regularization techniques like Lasso and Ridge modify the ordinary least squares objective function by adding a penalty proportional to the magnitude of the coefficients

Ans-8. D. SMOTE
SMOTE is a technique used to address class imbalance by oversampling the minority class. It works by generating synthetic examples in the feature space, creating new instances that are combinations of existing minority class instances. This helps to balance the class distribution and improve the performance of models, especially in classification tasks.

Ans-9. a. TRP and FRP
The AUC Receiver Operator Characteristic (ROC) curve uses A) TPR (True Positive Rate) and FPR (False Positive Rate) to plot the ROC curve and calculate the AUCROC.

Ans-10. B. False
In the AUC Receiver Operator Characteristic (AUCROC) curve the better model area under the curve should be less, therefore, for a better model, you would typically want the AUCROC to be closer to 1, not less. Higher AUCROC values represent better model performance in terms of distinguishing between the positive and negative classes.

Ans-11. A) Constructing a bag of words from an email

Bag of Words **(**BoW) is a technique used in natural language processing (NLP) for feature extraction from text data. It involves representing text data as a collection (bag) of words, ignoring grammar and word order, but keeping the multiplicity (frequency) of words.

Ans-12. Option A and B are correct.

in the context of the Normal Equation used for computing the coefficients of Linear Regression:

A) **We don't have to choose the learning rate**:

The Normal Equation is a closed-form solution that directly computes the optimal coefficients of the linear regression model without the need to manually select a learning rate, unlike gradient descent.

B) **It becomes slow when the number of features is very large**:

The Normal Equation involves computing the inverse of a matrix, which has a computational complexity of $O(n^3)$, As 'n' increases, the computation can become slow and memory-intensive.

Ans-13. Regularization in machine learning serves to address the balance between model complexity and performance on unseen data. Overfitting occurs when a model learns noise and peculiarities specific to the training data, leading to poor performance on new data. To mitigate this issue, regularization introduces a penalty to the model's optimization objective, discouraging overly complex models that fit the training data too closely.

There are primarily two types of regularization techniques widely used:

1. **L1 Regularization (Lasso):** Lasso adds a penalty equivalent to the absolute sum of the coefficients (L1 norm) to the model's loss function. This penalty encourages sparsity by shrinking some coefficients to zero, effectively performing feature selection. By reducing the number of features used in the model, Lasso helps simplify the model and improve interpretability.
2. **L2 Regularization (Ridge):** Ridge regression introduces a penalty equivalent to the squared sum of the coefficients (L2 norm) to the loss function. This penalty penalizes large coefficients and shrinks them towards zero, but typically does not eliminate them. Ridge regularization is effective in reducing the variance of the model and improving its stability, especially when dealing with multicollinearity among features.

Additionally, Elastic Net combines L1 and L2 penalties, offering a balance between feature selection (like Lasso) and regularization (like Ridge). This hybrid approach is useful in situations where there are multiple correlated predictors.

In summary, regularization techniques are crucial for improving the generalization of machine learning models by controlling model complexity and mitigating overfitting. They enhance model robustness and reliability, leading to better performance on new, unseen data while maintaining interpretability through feature selection and coefficient shrinkage.

Ans-14. Regularization techniques are not standalone algorithms themselves but are applied within algorithms to control model complexity and prevent overfitting. Here are some common algorithms in machine learning that incorporate regularization techniques:

1. **Linear Regression with Ridge (L2) Regularization**:
   - In linear regression, Ridge regression adds an L2 penalty term to the ordinary least squares objective function. This helps in reducing the coefficients' magnitude and thereby preventing overfitting.
2. **Linear Regression with Lasso (L1) Regularization**:
   - Lasso regression adds an L1 penalty term to the ordinary least squares objective function. It encourages sparsity by shrinking some coefficients to zero, effectively performing feature selection.
3. **Elastic Net Regression**:
   - Elastic Net combines both L1 (Lasso) and L2 (Ridge) penalties in the objective function. This hybrid approach provides a balance between Lasso and Ridge regularization, benefiting from both feature selection and regularization benefits.
4. **Logistic Regression with Regularization**:
   - Logistic regression can also incorporate L1 (Lasso) or L2 (Ridge) penalties to its loss function to prevent overfitting and improve generalization.
5. **Support Vector Machines (SVM)**:
   - SVMs can use L2 regularization in the form of soft margin SVMs, where a penalty is applied for misclassification errors to control the margin size and avoid overfitting.
6. **Neural Networks**:
   - Regularization techniques such as Dropout, L1/L2 regularization on weights, and early stopping are commonly used in training neural networks to prevent overfitting and improve generalization.
7. **Decision Trees with Pruning**:
   - Decision trees can be regularized by pruning techniques, where nodes or branches that provide little predictive power are removed to simplify the tree and prevent overfitting.

In summary, regularization techniques like L1 (Lasso), L2 (Ridge), and Elastic Net are integrated into various machine learning algorithms to improve their generalization performance and prevent overfitting.

Ans-15.

In linear regression, the term "error" refers to the discrepancy between the observed values of the dependent variable and the values predicted by the regression model. Here's a concise explanation suitable for a 5-mark response:

Linear regression is a statistical method used to model the relationship between a dependent variable (Y) and one or more independent variables (X). The fundamental equation of linear regression is: $Y = \beta 0 + \beta 1X + \epsilon$

The error term ϵ epsilonϵ embodies all factors influencing Y that are not accounted for by X. It is assumed to have a normal distribution with mean zero and constant variance. Minimizing the sum of squared residuals (differences between observed and predicted values) is essential in linear regression to ensure the model effectively captures the relationship between X and Y.

Understanding the error term ϵ is crucial for evaluating model accuracy, detecting outliers or influential data points, and validating the assumptions underlying linear regression. It quantifies the variability of data points around the fitted regression line, guiding the interpretation and reliability of the model's predictions.