**Find the Questions of test below of Machine Learning Theoretical Concepts**

**Linear Regression**

**1.What is the difference between simple linear regression and multiple linear regression?**

**Ans:** Single linear regression has one independent variable while multiple linear regression has multiple independent variables.

Single linear regression: y = ax + b

Multiple linear regression: y = a1x1 + a2x2 + a3x3….

**2.Explain the concept of the cost function in linear regression.**

**Ans:** The cost function is the measure of difference between the actual and predicted values.

IT shows how well the model fits the training data

MSE = sigma(y-yp)^2/n

**3.How do you interpret the coefficients in a linear regression model?**

**Ans:**

y = B0 + B1 X + e

Here, B0 is the intercept term, B1is the coefficient of the independent variable x and e represents the error term.(cost function)

In multiple linear regression with more than one independent variable, each coefficient represents the change in the dependent variable when the corresponding independent variable changes by one unit, while holding all other variables constant.

**4.What are the assumptions of linear regression?**

**Ans:**

1. Linearity: Relationship between variables is linear.

2. Independence: Observations are independent of each other.

3. Homoscedasticity: Constant variance of errors across all levels of the independent variables.

4. Normality: Residuals follow a normal distribution.

5. No multicollinearity: Independent variables are not highly correlated.

**Logistic Regression:**

**1.How does logistic regression differ from linear regression?**

Ans: logistic regression introduces non linearity in model, and is used to classify binary data, whereas linear regression is used when the data is linearly scattered.

Linear regression: y = b0+ b1x + b2x + c

Logistic regression: 1/(1+e^-x)

**2.Explain the sigmoid function and its role in logistic regression.**

**Ans :** The sigmoid function, is an activation function that maps any real-valued number to a value between 0 and 1. It has an S-shaped curve, which can convert values into probabilities.

In logistic regression, the sigmoid function is used to model the probability that a given observation belongs to a certain class (usually denoted as class 1). The output of the sigmoid function represents the probability of the event occurring

**3.What are the key performance metrics used to evaluate a logistic regression model?**

**Ans:** Accuracy, precision, recall, ROC, confusion matrix, specificity.

**4.How do you handle multicollinearity in logistic regression?**

**Ans:** Multicollinearity in logistic regression occurs when predictor variables are highly correlated with each other. To handle multicollinearity, we can use the techniques of regularization or PCA and selecting features that are least correlated.

**Naive Bayes:**

**1.What is the Naive Bayes algorithm based on?**

**And:** Naive bayes algo us based on the bayes probability theorem. It predicts the probability of one event happing given the probsblitliy of other event is given. It is called naive cause it assumes that the events are independent.

**2.Explain the concept of conditional probability in the context of Naive Bayes.**

**Ans:** Conditional probability, refers to the probability of an event occurring given that another event has already occurred. In the algorithm, it represents the likelihood of observing a particular feature value given a specific class label.

**3.What are the advantages and disadvantages of Naive Bayes?**

**Advantages:** fast prediction

Simple to implement

Efficient with high dimension data

**Disadvantages:**

Strong independence assumptions

Sensitive to imbalanced data

Cannot handle continuous features well

**4.How does Naive Bayes handle missing values and categorical features?**

**Ans;** During prediction, if a feature value is missing, Naive Bayes ignores that feature and calculates the probabilities based on the available features.

Naive Bayes naturally handles categorical features by treating each category as a separate feature.It calculates the conditional probabilities of each category given the class label.

**Decision Trees:**

**1.How does a decision tree make decisions?**

**ANs:** A decision tree makes decisions by recursively partitioning the feature space into subsets, based on the values of predictor variables, in order to classify or predict the target variable. It starts from the root node and splits the data into subsets using a selected feature and a corresponding threshold value, aiming to maximize information gain or minimize impurity.

**2.What are the main criteria for splitting nodes in a decision tree?**

**Ans:** The main criteria for splitting nodes in a decision tree include:

Gini Impurity: Gini impurity measures the likelihood of an incorrect classification of a randomly chosen element. It ranges from 0 (pure node) to 0.5 (maximum impurity). The split that minimizes the weighted average of the Gini impurity for child nodes is chosen.

Information Gain (Entropy): Information gain quantifies the reduction in entropy (uncertainty) achieved by splitting the data based on a particular feature. Entropy measures the randomness or disorder in a dataset. The feature that maximizes the information gain is selected for splitting.

Gain Ratio: Gain ratio is similar to information gain but adjusts for the number of branches that the split creates. It aims to avoid bias towards attributes with a large number of values. The feature that maximizes the gain ratio is chosen for splitting.

**3.How do decision trees handle categorical variables?**

**ANS:** Decision trees handle categorical variables by splitting the data based on the categories of the variable.

Binary Splitting: For binary categorical variables (e.g., yes/no, true/false), decision trees can split the data into two branches based on the presence or absence of the category.

One-Hot Encoding: Before building the decision tree, categorical variables are often encoded into binary dummy variables using one-hot encoding. Each category is represented by a binary variable (0 or 1), indicating the presence or absence of that category.

**4.What are some common techniques to prevent overfitting in decision trees?**

**Ans:** Pruning: Pruning involves removing parts of the tree that do not provide significant predictive power on the validation set. Pre-pruning stops the tree from growing beyond a certain depth or complexity, while post-pruning involves growing the tree fully and then removing branches that do not improve performance on the validation set.

Limiting Tree Depth: Restricting the maximum depth of the tree prevents it from becoming too complex and fitting the noise in the training data. This helps in generalizing better to unseen data.

**Support Vector Machines (SVM):**

**1.What is the basic idea behind SVM?**

**Ans:** The main idea behind Support Vector Machines (SVM) is to draw a line or boundary between different groups of points in a way that maximizes the gap between them. It's like finding the best road that keeps the groups farthest apart. This helps in making better predictions, especially when the groups are very close together

**2.Explain the concepts of margin and support vectors in SVM.**

**Ans:** the margin is the gap between the decision boundary (hyperplane) and the nearest data points from each class. It represents the separation between the different classes in the dataset. The goal of SVM is to find the hyperplane that maximizes this margin, meaning it tries to find the widest gap between the classes, which helps improve the model's ability to classify new data accurately.

Support vectors are the data points that lie closest to the decision boundary and directly influence the position and orientation of the hyperplane. These points are the most critical for defining the margin because if any support vector were to move, it would change the position of the decision boundary. In other words, support vectors are the "anchors" or "cornerstones" of the SVM model, as they determine the shape and position of the separating hyperplane.

**3.What are the different kernel functions used in SVM, and when would you use each?**

**ANs:**  kernels in SVM are:

Linear Kernel: Suitable for linearly separable data or when the number of features is large compared to the number of samples. It works well for simple classification tasks with linear decision boundaries.

Polynomial Kernel: Useful for capturing non-linear relationships between features. It introduces polynomial terms to the decision function, allowing SVM to fit more complex decision boundaries.

Radial Basis Function (RBF) Kernel: Most commonly used kernel for SVM. It is versatile and effective in capturing non-linear relationships. RBF kernel maps data into a high-dimensional space using Gaussian radial basis functions, enabling SVM to find complex decision boundaries.

Sigmoid Kernel: Suitable for binary classification problems where the relationship between features is similar to a sigmoid function. It can be useful for neural network-like architectures but is less commonly used compared to linear, polynomial, and RBF kernels.

Custom Kernels: In addition to these standard kernels, custom kernels can be defined based on domain knowledge or specific requirements of the problem.

**4.How does SVM handle outliers?**

**Ans:** use a soft-margin SVM, which allows for some misclassification to account for outliers while still aiming to maximize the margin.

Additionally, outlier detection techniques can be used in preprocessing to identify and remove or downweight outliers before training the SVM model.