Descriptive Qs

UNIT – I

1. Define Machine Learning and explain the main categories of Machine Learning algorithms. Provide an example of an application for each category
2. You are given a dataset that represents the relationship between a person's years of experience (feature) and their salary (target). Explain how linear regression can be used to model and predict the salaries based on years of experience. Describe the linear regression model and the key components involved.
3. Define regression analysis in the context of Machine Learning. Elaborate on the difference between simple linear regression and multiple linear regression. Provide examples of scenarios where each type of regression might be applicable.
4. You are provided with a dataset containing information about house prices. Discuss how estimation functions, specifically the least squares method, can be employed to develop a regression model that predicts house prices based on features like square footage, number of bedrooms, and location.

UNIT -II

1. Define classification in the context of Machine Learning. Explain the difference between binary and multiclass classification with examples of each.
2. You have built a classification model to distinguish between cats and dogs using a dataset of images. The model's performance needs to be evaluated. Describe how accuracy, precision, recall, and F1-score are calculated. Highlight their significance in the evaluation process and provide scenarios where one metric might be more relevant than others
3. Explain the fundamental concept of classification in Machine Learning. Compare and contrast supervised and unsupervised classification algorithms, providing an example of each type.
4. You have constructed a spam email classifier using a labeled dataset. The model's performance must be assessed. Detail the concepts of accuracy, precision, recall, and F1-score. Emphasize the importance of these metrics in evaluating the classifier's effectiveness, and present scenarios where each metric is more relevant.

UNIT- III

1. Explain what resampling methods are in Machine Learning and why they are used. Compare and contrast bootstrapping and cross-validation techniques, highlighting their purposes and how they address overfitting.
2. Describe the concept of regularization in linear regression. What problem does it solve? Explain L1 regularization (Lasso) and L2 regularization (Ridge), including how they modify the linear regression cost function. Elaborate on the effects of the regularization parameters on the model's complexity and the importance of feature selection.
3. Define resampling techniques in Machine Learning and elucidate their significance. Contrast the process of k-fold cross-validation with leave-one-out cross-validation. Describe how these techniques help in managing issues like overfitting.
4. Clarify the concept of regularization and its purpose in preventing overfitting in models. Dive into L1 regularization (Lasso) and L2 regularization (Ridge) in linear regression. Expound on how these methods modify the linear regression cost function. Discuss the impact of different regularization strengths on model complexity and the necessity of feature selection.

|  |
| --- |
| UNIT -I Describe the fundamental concepts of Machine Learning. Explain the main categories of Machine Learning algorithms and provide an example of a real-world application for each category. Discuss how these categories differ in terms of data requirements, learning approach, and problem types they address ? |
| UNIT- I Explain the process of linear regression in Machine Learning. Define the concepts of dependent and independent variables, and detail how the least squares method is used to estimate the coefficients of a linear regression model. Discuss the assumptions underlying linear regression and their significance in model interpretation and validity? |
| UNIT- II Define classification in the context of Machine Learning. Compare and contrast binary and multiclass classification, providing examples of each. Explain the concepts of true positive, true negative, false positive, and false negative. Discuss the importance of each of these metrics in evaluating a classification model's performance.? |
| UNIT- II Discuss the challenges of handling imbalanced classes in classification tasks. Describe the techniques or strategies that can be employed to address class imbalance and improve the performance of the model. Provide examples of situations where class imbalance might occur and explain how these techniques can mitigate bias towards the majority class? |
| UNIT- III Define resampling methods in Machine Learning and explain their purpose in model assessment. Compare and contrast bootstrapping and cross-validation techniques, highlighting their advantages and differences in addressing overfitting. Discuss how these techniques can provide insights into a model's performance on unseen data? |
| UNIT- III Discuss the concept of regularization in Machine Learning. Explain the trade-off between model complexity and model performance. Describe L1 regularization (Lasso) and L2 regularization (Ridge) in the context of linear regression. Explain how these techniques modify the cost function and how the regularization parameter affects the coefficients and model complexity? |

**Open Book :**

UNIT- I

Machine Learning algorithms have shown great potential in healthcare for early disease detection. Consider a scenario where you are working with a dataset containing medical records of patients. The goal is to predict the likelihood of a patient developing diabetes within the next five years. Apply linear regression to build a predictive model based on relevant features such as age, BMI, and blood pressure. Discuss how you would address potential challenges in the dataset and model interpretation, and explain how your model's predictions could assist healthcare professionals

Urban planning and development require accurate predictions of property values. Assume you are working with a dataset that contains information about housing prices in a city, including features like square footage, number of bedrooms, and distance from city center. Develop a multiple linear regression model to predict property values. Detail the process of data preprocessing, including handling missing values and categorical variables. Explain how you would use the coefficients from the model to provide insights to city planners for informed decisions

UNIT- II

E-commerce platforms often face the challenge of fraud detection. Suppose you are given a dataset containing transaction records, including features like transaction amount, location, and time. Your task is to build a fraud detection model using binary classification. Apply suitable classification algorithms and emphasize the importance of precision in this context. Describe how you would evaluate the model's performance and discuss the potential consequences of false positives and false negatives

A medical laboratory aims to automate the diagnosis of certain diseases from medical images. Consider a dataset of X-ray images for diagnosing respiratory conditions. Apply a deep learning approach to build a classification model capable of identifying different respiratory conditions. Choose appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Describe how the model's predictions can assist medical professionals in making accurate diagnoses and the potential ethical considerations in using AI for medical decisions

UNIT- III

Real estate pricing models can be sensitive to the choice of features and model complexity. Consider a dataset with attributes like property size, location, and amenities. Build a linear regression model to predict property prices and introduce regularization techniques (L1 and L2). Explain how regularization can prevent overfitting and justify your choice between Lasso and Ridge based on the dataset's characteristics. Discuss how regularization terms impact the coefficients and the interpretability of the model.

The performance of a classification model often depends on the quality and quantity of labeled data. Imagine you are developing a sentiment analysis system for customer reviews. Apply k-fold cross-validation to evaluate the model's accuracy and generalize its performance. Discuss the benefits of cross-validation in handling the limited availability of labeled data and explain how this technique helps in estimating the model's performance on new and unseen data.

**PATTERN**

Concept definition

Concept explanation with steps

consider an example data set or given dataset and apply the concept on the dataset

evaluate the model by applying performance matrix

report the results