UNIT-4

Linear methods for Regression:
Introduction, Linear Regression models, Least Squares, Multiple Regression.
Linear methods for Classification: Introduction, Linear discriminative analysis, Logistic Regression.

Linear methods for Regression:

Linear Regression is a widely used statistical method for modeling the relationship between a dependent variable (often denoted as Y) and one or more independent variables (often denoted as X). The basic idea behind linear Regression is to find the best fitting straight line that represents the relationship between Y and X. There are several linear methods for Regression, including simple linear Regression, multiple linear Regression, and polynomial Regression.

1. Simple Linear Regression: Simple linear Regression is used when only one independent variable (X) is used to predict the dependent variable (Y). For example, if we want to predict the weight of a person based on their height, we can use simple linear Regression with height as the independent variable and weight as the dependent variable. The equation for simple linear Regression can be written as Y = a + bX, where a is the intercept, b is the slope, and X is the independent variable.

2. Multiple Linear Regression: Multiple linear Regression is used when two or more independent variables (X1, X2, etc.) are used to predict the dependent variable (Y). For example, if we want to predict the salary of an employee based on their education level, years of experience, and job title, we can use multiple linear Regression with education level, years of experience, and job title as the independent variables and salary as the dependent variable. The equation for multiple linear Regression can be written as Y = a + b1X1 + b2X2 + ... + bnXn, where a is the intercept, b1, b2, etc. are the slopes for each independent variable, and X1, X2, etc. are the independent variables.

3. Polynomial Regression: Polynomial Regression is used when the relationship between the independent and dependent variables is not linear but can be approximated by a polynomial function. The equation for polynomial Regression can be written as Y = a + b1X + b2X^2 + ... + bnx^n, where X is the independent variable, b1, b2, etc., are the coefficients for each degree of X, and n is the degree of the polynomial. For example, suppose we want to predict the sales of a product based on the advertising budget. In that case, we can use polynomial Regression with the advertising budget as the independent variable and sales as the dependent variable.

Here are some examples of linear Regression in different fields:

1. Finance: Linear Regression can be used to model the relationship between a company's stock price and various economic factors such as interest rates, inflation, and GDP growth. By analyzing the historical data, analysts can predict how the stock price will likely change based on changes in these economic factors.
2. Marketing: Linear Regression can be used to predict sales based on advertising spending. By analyzing historical data, marketers can determine the optimal amount of advertising spending to maximize sales.
3. Healthcare: Linear Regression can be used to model the relationship between a patient's medical history and their risk of developing a particular disease. By analyzing the historical data, doctors can identify the risk factors and take preventive measures to reduce the risk of the disease.
4. Environmental Science: Linear Regression can be used to model the relationship between various environmental factors, such as temperature, rainfall, and air pollution, and their impact on ecosystems. By analyzing the historical data, scientists can predict how changes in these environmental factors will affect the ecosystem in the future.
5. Engineering: Linear Regression can be used to predict the performance of a product based on various design factors such as materials, dimensions, and manufacturing processes. By analyzing the historical data, engineers can optimize the product's design to improve its performance and reduce costs.

Linear Methods for Classification:

Linear classification methods are a family of techniques that involve using linear decision boundaries to classify data points into different classes. Here are some commonly used linear methods for classification:

1. Logistic Regression: Logistic Regression is a popular linear method for binary classification, which involves modeling the probability of a data point belonging to a particular class. The decision boundary is linear, and the algorithm learns the weights of the linear function using maximum likelihood estimation.
2. Linear discriminant analysis (LDA): LDA is a linear method that finds the linear combination of features that maximally separates the classes. It projects the data onto a lower-dimensional space, where the decision boundary is a hyperplane. LDA can be used for both binary and multi-class classification.

3. Support vector machines (SVMs): SVMs are a powerful linear classifiers that work by finding the hyperplane that maximally separates the classes. SVMs are particularly useful when the data is not linearly separable, as they can use kernel functions to transform the data into a higher-dimensional space where the classes are separable.
4. Perceptron: The perceptron is a simple linear classifier that works by iteratively updating the weights of a linear function until the training data is correctly classified. The decision boundary is a hyperplane, and the algorithm is particularly useful for binary classification problems.

Linear classification methods are widely used in many fields, including finance, healthcare, and image recognition. For example, logistic Regression is commonly used in credit scoring to predict the likelihood of a borrower defaulting on a loan. In contrast, SVMs are used in image recognition to classify images into different categories.

Linear classification is a technique used to classify data points into different classes based on linear decision boundaries. Linear classifiers work by finding the best hyperplane or line that separates the different classes of data points. This approach assumes that the data points can be separated using a linear boundary.

Linear Discriminative Analysis:

Linear discriminative analysis (LDA) is a linear classification method used to find the linear combination of features that best separates the classes of data points. LDA aims to find the hyperplane that best separates the data points from different classes. LDA is used for both binary and multi-class classification problems.

For example, suppose we have a dataset containing the heights and weights of male and female students. We can use LDA to find the hyperplane that best separates the male and female students based on their heights and weights. Once we have this hyperplane, we can use it to classify new data points as male or female.

Logistic Regression:

Logistic Regression is a popular linear classification method used for binary classification problems. In logistic Regression, we model the probability of a data point belonging to a particular class as a function of its features. The decision boundary is a linear function that separates the data points into two classes.

For example, suppose we have a dataset containing information about customers who either buy or do not buy a product. We can use logistic Regression to predict

the probability of a customer buying the product based on age, gender, income, and other features. Once we have the probability, we can use a threshold value to classify the customer as either a buyer or a non-buyer.

Logistic Regression is widely used in many fields, including healthcare, finance, and marketing. For example, logistic Regression is used in healthcare to predict the likelihood of a patient developing a particular disease based on their medical history and other risk factors. In finance, logistic Regression is used to predict the likelihood of a borrower defaulting on a loan based on their credit history and other financial indicators.