222CS3113

..

Write short notes on PCA and LDA. Distinguish between Bayesian Parameter Estimation and Maximum Likelihood Estimate.

PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) are both popular dimensionality reduction techniques used in machine learning and data analysis.

PCA is an unsupervised technique used for feature extraction and data compression. It aims to identify the underlying patterns and relationships in the data by finding a set of orthogonal axes (principal components) that account for the largest amount of variation in the data. PCA can be used to reduce the dimensionality of a dataset by projecting the data onto a lower-dimensional space defined by the principal components.

LDA, on the other hand, is a supervised technique used for feature extraction and classification. It aims to find a set of linear combinations of the input features that maximize the separation between different classes. LDA can be used to reduce the dimensionality of a dataset while preserving the class-discriminatory information.

The main difference between PCA and LDA is that PCA is an unsupervised technique that focuses on finding the directions of maximum variance in the data, while LDA is a supervised technique that focuses on finding the directions that best separate the different classes in the data.

Bayesian Parameter Estimation and Maximum Likelihood Estimate (MLE) are two common methods used for estimating the parameters of a statistical model. Bayesian parameter estimation involves using Bayes' theorem to update the prior belief about the parameters based on new data. This approach produces a posterior distribution over the parameters that takes into account both the prior belief and the likelihood of the data. MLE, on the other hand, involves finding the parameter values that maximize the likelihood of the data. This approach does not take into account any prior beliefs about the parameters.

The main difference between these two methods is that Bayesian parameter estimation incorporates prior knowledge or beliefs about the parameters, while MLE does not. Bayesian parameter estimation is useful when prior knowledge is available or when a regularizing effect is desired, while MLE is useful when no prior knowledge is available or when a more parsimonious model is desired.

With respect to linear regression, consider the given information and answer the given questions:
The sales of a company (in million dollars) for each year are shown in the table below:
x (year), 2015, 2016, 2017, 2018, 2019
y (sales),12 ,19, 29 ,37, 45
i) Find the least square regression line $y = a x + b$**.**
ii) Use the least squares regression line as a model to estimate the sales of the company in 2022.

i) To find the least squares regression line, we need to compute the slope (a) and y-intercept (b) of the line that minimizes the sum of squared errors between the predicted and actual sales values. The formulas for the slope and y-intercept are:

$a = (n\Sigma(xy) - \Sigma x\Sigma y) / (n\Sigma(x^2) - (\Sigma x)^2)$

$b = (\Sigma y - a\Sigma x) / n$

where n is the number of data points, Σ represents the sum of the indicated values, and x and y represent the year and sales values, respectively.

Using these formulas with the given data, we get:

$n = 5$
$\Sigma x = 100$
$\Sigma y = 142$
$\Sigma xy = 4625$

$\Sigma(x^2) = 1015$

$a = (54625 - 100142) / (51015 - 100^2) \approx 8.2$

$b = (142 - 8.2100) / 5 \approx -17.4$

Therefore, the least squares regression line is:

$y = 8.2x - 17.4$

ii) To estimate the sales of the company in 2022 using the least squares regression line, we need to substitute x=2022 into the equation for y:

$y = 8.2(2022) - 17.4 \approx 16429.6$

Therefore, the model predicts that the sales of the company in 2022 will be approximately \$16,429.6 million. It's important to note that this is just an estimate based on the linear regression model, and actual sales may differ depending on various factors.

Identify the following problem with given four training points: (1, 1, -), (1,0, +), (0, 1, +), (0, 0, -). Check the data are linearly separable or not. If not then, transform the data from input space to feature space using following transformation function φ. Find the maximum margin classifier (linear decision boundary) in the transformed space.
φ = (1,√2x_1,√2x_2,√2x_1x_2, (x_1)^2, (x_2)^2)

The given problem is a binary classification problem, where the goal is to learn a decision boundary that can separate the two classes labeled as "+" and "-". Each data point has three features, two input features (x1, x2) and a target label (+ or -).

To check whether the data is linearly separable, we can try to draw a decision boundary that can perfectly separate the two classes. However, in this case, it is not possible to draw a straight line that can separate the two classes. We can see this by plotting the four training points on a graph. The two classes are mixed, and there is no clear boundary between them.

To transform the data from input space to feature space, we can use the given transformation function φ. This transformation maps the input space to a higher-dimensional feature space where the data may become separable.

The transformed feature space is:

φ = (1, √2x1, √2x2, √2x1x2, x1^2, x2^2)

Using this transformation, we can map each of the four training points to the corresponding feature space points:

(1, 1, -) → (1, √2, √2, √2, 1, 1)
(1, 0, +) → (1, √2, 0, 0, 1, 0)
(0, 1, +) → (1, 0, √2, 0, 0, 1)

(0, 0, -) → (1, 0, 0, 0, 0, 0)

Now, we can try to find a maximum margin classifier (linear decision boundary) in the transformed feature space. The maximum margin classifier is the hyperplane that maximizes the distance between the two closest points from different classes (also called support vectors).

To find the maximum margin classifier, we can use a linear classification algorithm such as support vector machine (SVM). The SVM tries to find a hyperplane in the feature space that separates the two classes and maximizes the margin between the hyperplane and the support vectors.

Using an SVM with a linear kernel, we can find the maximum margin classifier in the transformed feature space. The decision boundary in the feature space is:

w0 + w1√2x1 + w2√2x2 + w3√2x1x2 + w4(x1^2) + w5(x2^2) = 0

where w0, w1, w2, w3, w4, w5 are the weights learned by the SVM.

The maximum margin classifier in the input space is the boundary that corresponds to the above equation. It is a nonlinear decision boundary in the input space that separates the two classes.

Note that while the transformed data is linearly separable, the original data is not. The transformation function φ has mapped the original data to a higher-dimensional space where it becomes linearly separable.

Define Linear Regression Model in Mathematical formation with proper explanation of each term. Last year, six randomly selected students took a Mathematics aptitude test before they began their Machine Learning course. Find the linear regression equation which best predicts Machine Learning performance, based on Mathematics aptitude scores? In the table, the *xi* column shows scores on the aptitude test. Similarly, the *yi* column shows machine learning grades. Calculate the absolute error on sample 4. Define Multiple regression model and How do you estimate the model parameters

Student, *Mathematics Score* **(***xi***) ,***ML Grade* **(***yi***)**
1, 85, 90
2 ,80, 95
3 ,75, 80
4 ,65 ,70
5 ,60 ,65
6, 55, 75

Linear Regression is a statistical technique that is used to model the relationship between a dependent variable and one or more independent variables. It tries to find the best-fit straight line equation that describes the linear relationship between the variables.

The mathematical equation for linear regression with a single independent variable is given by:

y = b0 + b1 * x + e

where,

y is the dependent variable (the variable we want to predict)
x is the independent variable (the variable we use to predict the dependent variable)
b0 is the y-intercept (the value of y when x is 0)
b1 is the regression coefficient (the slope of the line)

e is the error term (the difference between the predicted value and the actual value)

In the given problem, we have one independent variable (Mathematics aptitude scores) and one dependent variable (Machine Learning grades). To find the linear regression equation, we need to calculate the regression coefficient (b1) and y-intercept (b0) using the following formulas:

b1 = ∑((xi - x_mean) * (yi - y_mean)) / ∑((xi - x_mean)^2)

b0 = y_mean - b1 * x_mean

where,

xi is the ith value of the independent variable
yi is the ith value of the dependent variable
x_mean is the mean of the independent variable

y_mean is the mean of the dependent variable

Using the given data, we can calculate the regression coefficient (b1) and y-intercept (b0) as follows:

x_mean = (85 + 80 + 75 + 65 + 60 + 55) / 6 = 70

y_mean = (90 + 95 + 80 + 70 + 65 + 75) / 6 = 78.33

b1 = [((85 - 70) * (90 - 78.33)) + ((80 - 70) * (95 - 78.33)) + ((75 - 70) * (80 - 78.33)) + ((65 - 70) * (70 - 78.33)) + ((60 - 70) * (65 - 78.33)) + ((55 - 70) * (75 - 78.33))] / [((85 - 70)^2) + ((80 - 70)^2) + ((75 - 70)^2) + ((65 - 70)^2) + ((60 - 70)^2) + ((55 - 70)^2)] = 0.77

b0 = 78.33 - (0.77 * 70) = 26.23

Therefore, the linear regression equation is:

y = 0.77x + 26.23

To calculate the absolute error on sample 4, we can substitute the value of x as 65 in the above equation to get the predicted value of y and then subtract it from the actual value of y for sample 4.

Absolute error = |predicted value of y - actual value of y| = |(0.77 * 65 + 26.23) - 70| = 1.37

Multiple regression is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. The equation for multiple regression is similar to that of linear regression with a single independent variable, but with additional regression coefficients for each independent variable.

y = b0 + b1 * x1 + b2 * x2 + ... + bn * xn + e

where,

y is the dependent variable
x1, x2, ..., xn are the independent variables

b0 is the y-intercept