Theory of LDA

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique commonly used in machine learning and statistics for classification tasks.

Linear Discriminant Analysis aims to find a linear combination of features that characterizes or separates classes in the dataset. It achieves this by maximizing the separation between classes while minimizing the variation within each class.

**Example: Credit Risk Assessment**

**Objective:**

The objective is to develop a model that can accurately classify loan applicants as either low-risk or high-risk based on their financial attributes and credit history.

**Features:**

Features could include various financial attributes and credit history information of loan applicants, such as income, debt-to-income ratio, credit score, number of open accounts, number of delinquencies, etc.

Each feature provides valuable information about the creditworthiness of the applicant.

**Application of LDA:**

**Data Collection:**

Gather a dataset containing information about past loan applicants, including their financial attributes, credit history, and the outcome of their loan applications (approved or denied).

**Feature Extraction:**

Extract relevant features from the dataset, focusing on financial attributes and credit history variables.

Normalize or standardize the features to ensure that they are on the same scale.

**Dimensionality Reduction with LDA:**

Apply Linear Discriminant Analysis (LDA) to the dataset to reduce the dimensionality of the feature space while maximizing the separation between low-risk and high-risk loan applicants.

LDA will find a linear combination of features that best discriminates between the two classes of loan applicants.

**Model Training and Testing:**

Train a classification model (e.g., Logistic Regression, Decision Trees) using the reduced feature space obtained from LDA.

Split the dataset into training and testing sets for model evaluation.

**Model Evaluation:**

Evaluate the trained model on the testing set to assess its performance in correctly classifying loan applicants as low-risk or high-risk.

Metrics such as accuracy, precision, recall, and F1-score are calculated to measure the model's effectiveness