# Linear Discriminant Analysis

Supervised Dimensionality Reduction.

# Intro

Linear Discriminant Analysis (LDA) is a commonly used dimensionality reduction technique. However, despite the similarities to Principal Component Analysis (PCA), it differs in one crucial aspect.

Instead of finding new axes (dimensions) that maximize the variation in the data, it focuses on **maximizing the separability** among the known categories (classes) in the target variable.

## Applying Principal Component Analysis (PCA) concepts:

Since PCA is an unsupervised Machine learning technique, we do not need to provide a target variable with category labels. This means that PCA does not care about whether the apartment is part of a “cheap” or “expensive” category.

# The goal of PCA is to capture the maximum amount of variance, which is achieved by the algorithm finding a line that minimizes the distances from data points to that line.

Interestingly, this is equivalent to maximizing **the spread of data point projections on that same line, which is why we can** capture the maximum amount of variance.

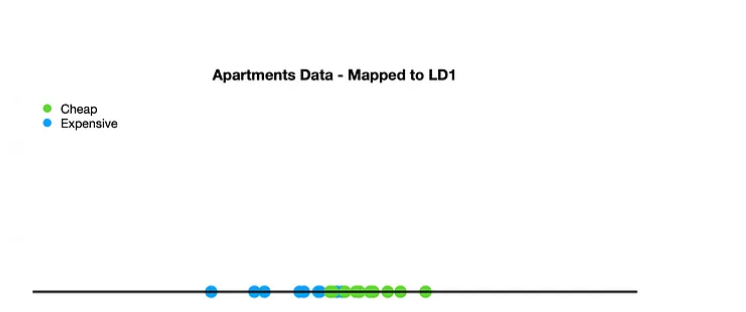
As you can see from the graph above, we have done a good job finding a new axis that preserves most of the variance, enabling us to go down from two dimensions to one. This is what it looks like after we map the data to our new single dimension (PC1):



You can clearly see that, in this scenario, we were able to preserve most of the variance but still lost some information that would have been useful when trying to separate the two categories (classes). If you see the two classes using PCA are overlapped.

## Applying Linear Discriminant Analysis (LDA)

Now let’s go back to our original example data and apply LDA instead of PCA. Quick reminder, **the goal of LDA is to maximize the separability of the known categories in our target variable** (“cheap”, “expensive”) while at the same time reducing dimensions.



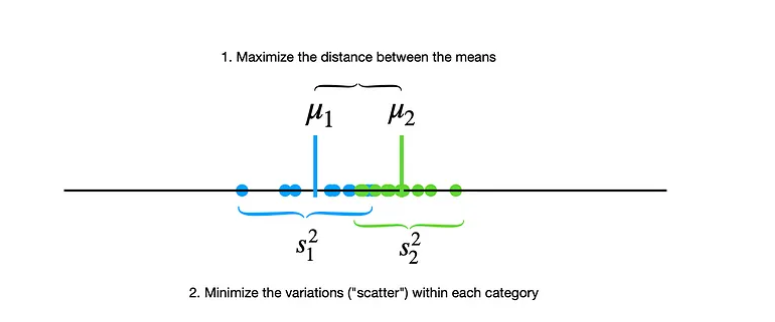
While the separation is imperfect, it is visibly better, with only a few overlapping observations from the two categories.

## How does LDA find the right axis?

Two key criteria are used when finding the “best” line for our new axis, which are considered simultaneously.

1. **Maximizing the distance (d²).**  
   **2 categories**— when you have two classes in your target variable, the distance refers to the difference between the mean (μ) of class 1 and the mean of class 2.  
   **More than 2 categories** — when you have three or more classes in your target variable, the algorithm first finds a central point to all of the data and then measures the distance from each category mean (μ) to that central point.

2. **Minimize the variation, also known as “scatter” in LDA (s²).**



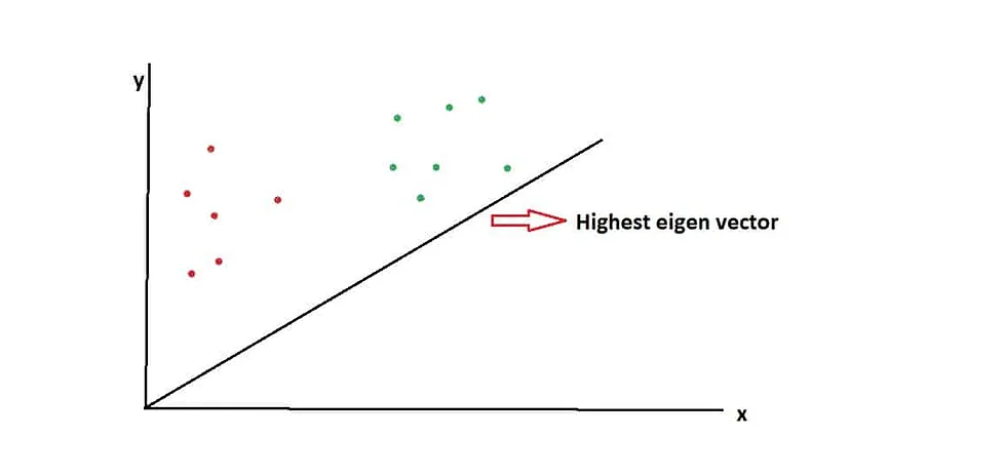
In practice, the calculations are typically performed with the help of either Singular Value Decomposition (SVD) or Eigenvalue decomposition. Both of these options are available in sklearn’s LDA implementation.

Linear Discriminant Analysis is a method of Dimensionality Reduction. The goal of LDA is to project a dataset onto a lower-dimensional space. It sounds similar to PCA. Right?

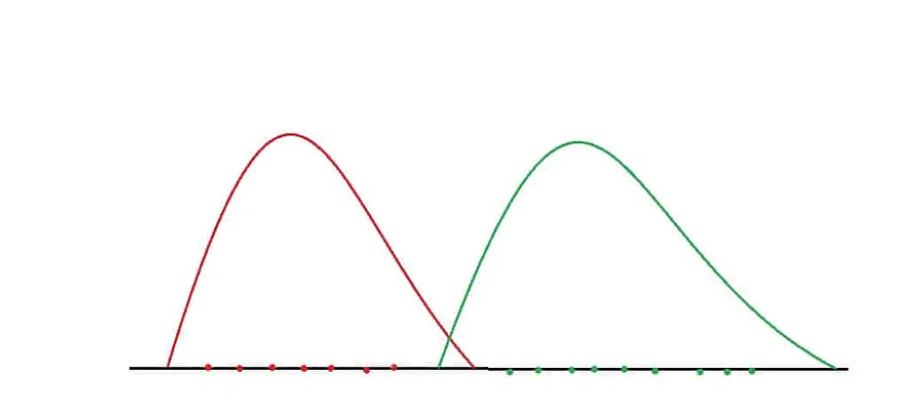
But LDA is different from PCA. Linear Discriminant Analysis finds the area that maximizes the separation between multiple classes. That is not done in PCA.

So, the definition of LDA is- **LDA project a feature space (N-dimensional data) onto a smaller subspace k( k<= n-1) while maintaining the class discrimination information.**

PCA is known as**Unsupervised**but LDA is **supervised** because of the relation to the dependent variable.



When data points are projected onto this vector, so the dimensionality is reduced as well as the discrimination between the classes is also maximized.



This line of one dimension maximizes the mean distance of the two classes.

# LDA mean focus is to maximize the separation between classes without taking into consideration the loss of variance. Meantime PCA does not care about the classes’ separation, it focuses on reducing the dimension of the data and keeping the maximum amount of variance, relevant data.

# When you know the labels of your data and you want to use a model for prediction LDA is better than PCA for dimensionalty reduction.

LDA is a great tool when we want to reduce the dimensionality of our data while keeping as much information relevant to our prediction target.

However, the direct comparison of PCA and LDA might not be completely fair because they are two different techniques meant for different purposes. For example, PCA is an **unsupervised** learning technique, while LDA falls under the **supervised** branch of ML.

Hence, please don’t take LDA superiority over PCA. Instead, please first assess the applicability of these algorithms to your unique case.