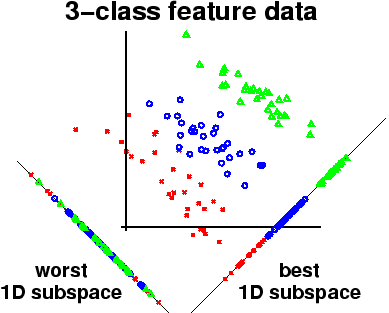
LINEAR DISCRIMINANT ANALYSIS

* LDA works when the measurements made on independent variables for each observation are continuous quantities.
* The use of Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition.
* Logistic regression is a classification algorithm traditionally limited to only two-class classification problems. If you have more than two classes then Linear Discriminant Analysis is the preferred linear classification technique.
* Mean as far as possible with variance as small as possible. Whereas PCA looks for greatest variance.
* If separating information is not in the means but in the variance then LDA would fail and PCA will give better result.
* If your data all belongs to the same class, then you might be interested more in [PCA (Principcal Component Analysis)](http://en.wikipedia.org/wiki/Principal_component_analysis), which gives you the most important directions for the data ranked in order of importance.

1. Quora :

Let us say you have data that is represented by 100 dimensional feature vectors and you have 100000 data points. You know/suspect that these data points belong to three different classes but you are not sure which features/combination of features are mostly affecting their separation. The data you have is too large to perform any reasonable computation in reasonable time. So you want to reduce these 100 dimensional feature vector to say 50 dimensional feature vector to allow you to learn the data more efficiently.  
  
Performing Principal Component Analysis to reduce the number of features (dimensions) would have given you which all features affected your data by computing their leading eigenvalues. But you are not satisfied, though you have obtained the 50 new features, they do not correctly distinguish the 3 classes as they were in the original data.  
  
You want to preserve as the difference between the classes as well while reducing the dimensions. You look for a better alternative, and it leads you to Linear Discriminant Analysis which reduces the number of features by also considering the inter-class separation between the classes.  
  
LDA just reduces the number of dimensions of the input feature vector by preserving the inter-class separation as present in the original feature vector.

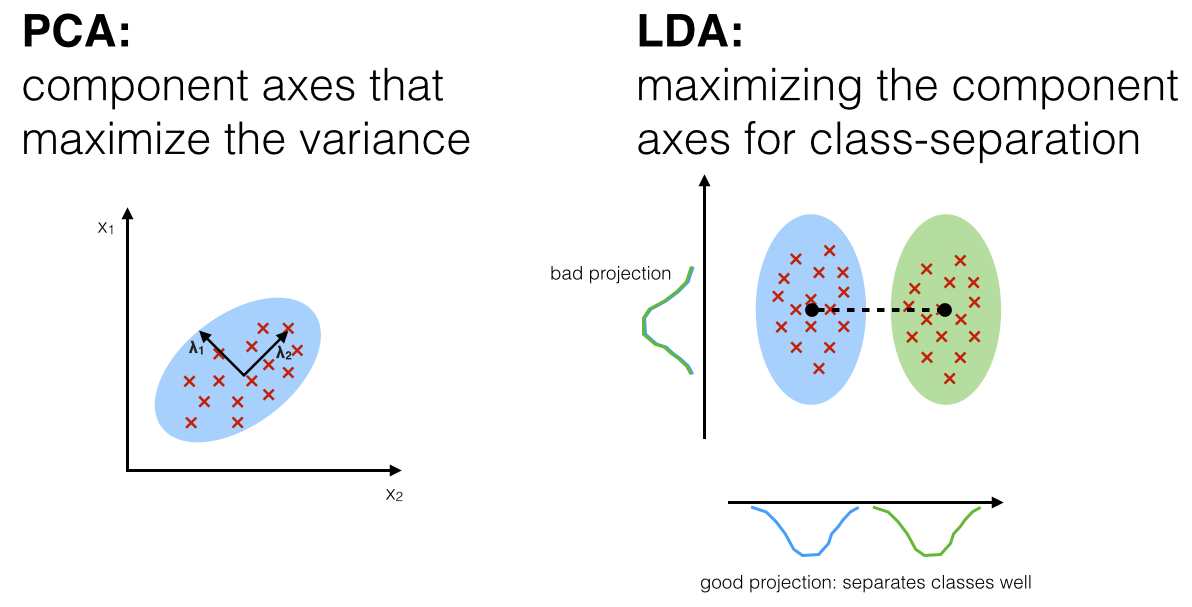


In the above diagram, a 3 dimensional input feature vector is reduced to 1 dimensional feature vector in the meantime preserving the differences among the classes.

1. **Principal Component Analysis vs. Linear Discriminant Analysis**

Both Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are linear transformation techniques that are commonly used for dimensionality reduction. PCA can be described as an “unsupervised” algorithm, since it “ignores” class labels and its goal is to find the directions (the so-called principal components) that maximize the variance in a dataset. In contrast to PCA, LDA is “supervised” and computes the directions (“linear discriminants”) that will represent the axes that that maximize the separation between multiple classes.

Although it might sound intuitive that LDA is superior to PCA for a multi-class classification task where the class labels are known, this might not always the case.  
For example, comparisons between classification accuracies for image recognition after using PCA or LDA show that PCA tends to outperform LDA if the number of samples per class is relatively small ([PCA vs. LDA](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=908974), A.M. Martinez et al., 2001). In practice, it is also not uncommon to use both LDA and PCA in combination: E.g., PCA for dimensionality reduction followed by an LDA.



### Normality assumptions

### It should be mentioned that LDA assumes normal distributed data, features that are statistically independent, and identical covariance matrices for every class. However, this only applies for LDA as classifier and LDA for dimensionality reduction can also work reasonably well if those assumptions are violated. And even for classification tasks LDA seems can be quite robust to the distribution of the data

1. **Formula :**

http://people.revoledu.com/kardi/tutorial/LDA/image/LDA_clip_image102.gif

Assign object k to group i that has maximum Fi.

Numerical Eg : http://people.revoledu.com/kardi/tutorial/LDA/Numerical%20Example.html