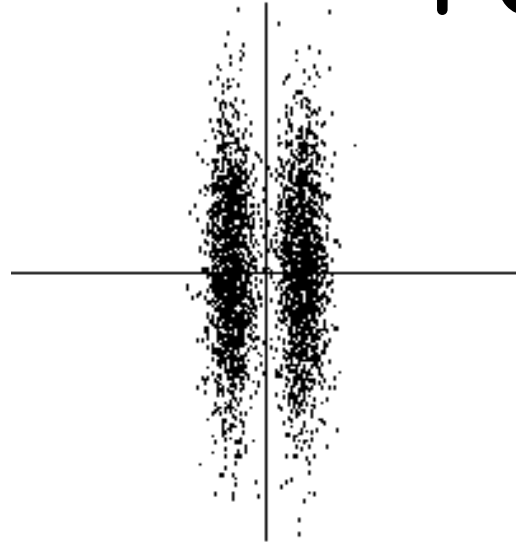


# Linear Discriminant Analysis (LDA)

Ahmet Sacan

PCA

# PCA

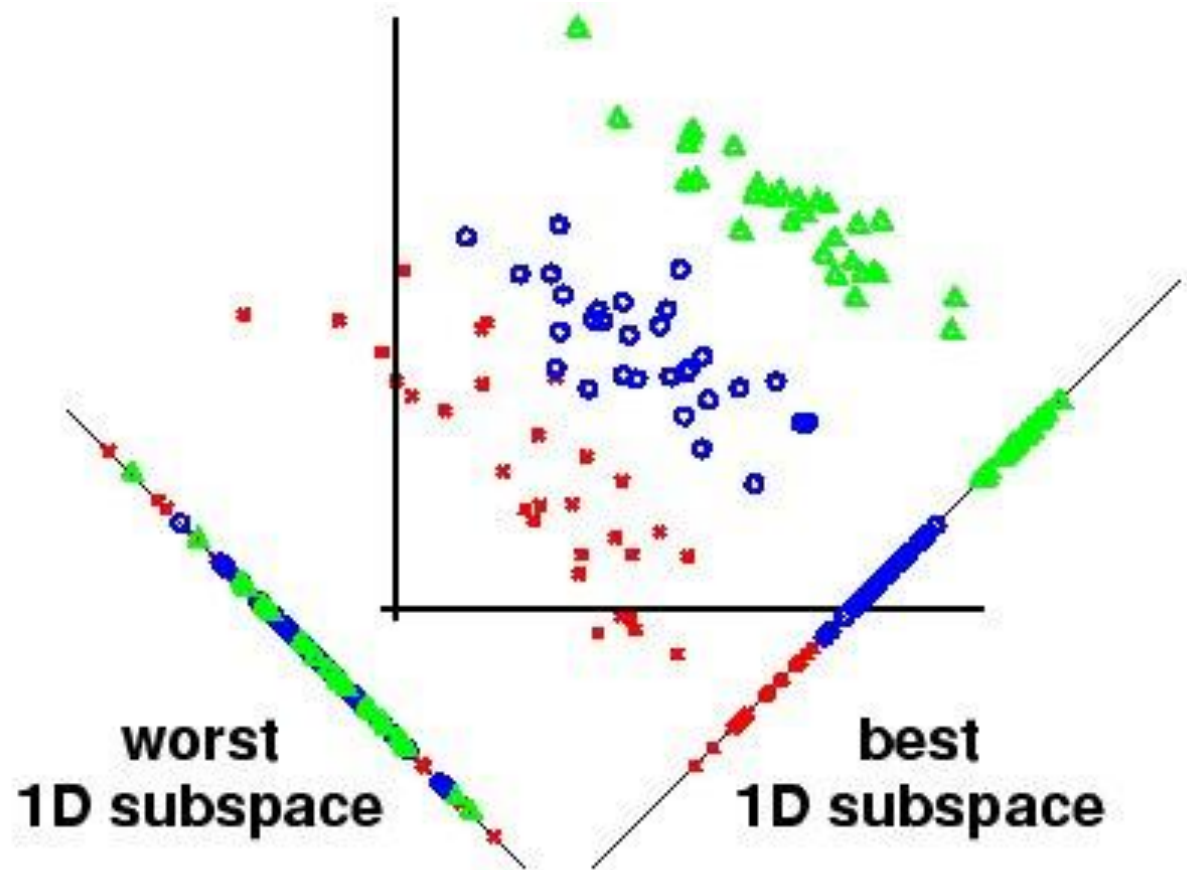


- The data is clearly divided into two clusters and each cluster has gaussian distribution
- PCA chooses the vertical axis as the principal component because of higher variance
  - Clustering information is lost
- Which projection would be more "interesting"?

# Projection based on Interestingness (Projection Pursuit)

- Specify the projection index function  $I(x, \alpha)$  that maximizes the interestingness of the projection
  - Measure of variation (PCA)
  - Departure from normality (negative entropy)
  - Class separability (LDA, Bhattacharyya)
  - ...
- Solve for  $\alpha$  by maximizing  $I(x, \alpha)$

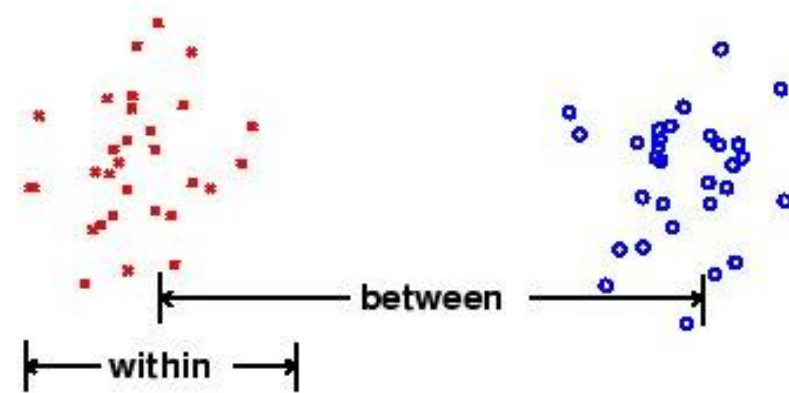
# LDA



# Within/between class scatter

- Within-class scatter ( $S_w$ )

$$\begin{aligned} S_w &= \sum_j \text{cov}_j \\ &= \sum_j (X_j - \mu_j)(X_j - \mu_j)^T \end{aligned}$$



- where  $j$  is one of the red/blue classes.
- Between-class scatter ( $S_b$ )

$$S_b = \sum_j (\mu_j - \mu_*)(\mu_j - \mu_*)^T$$

# LDA Criterion

- Maximize ratio of overall variance to within class variance:

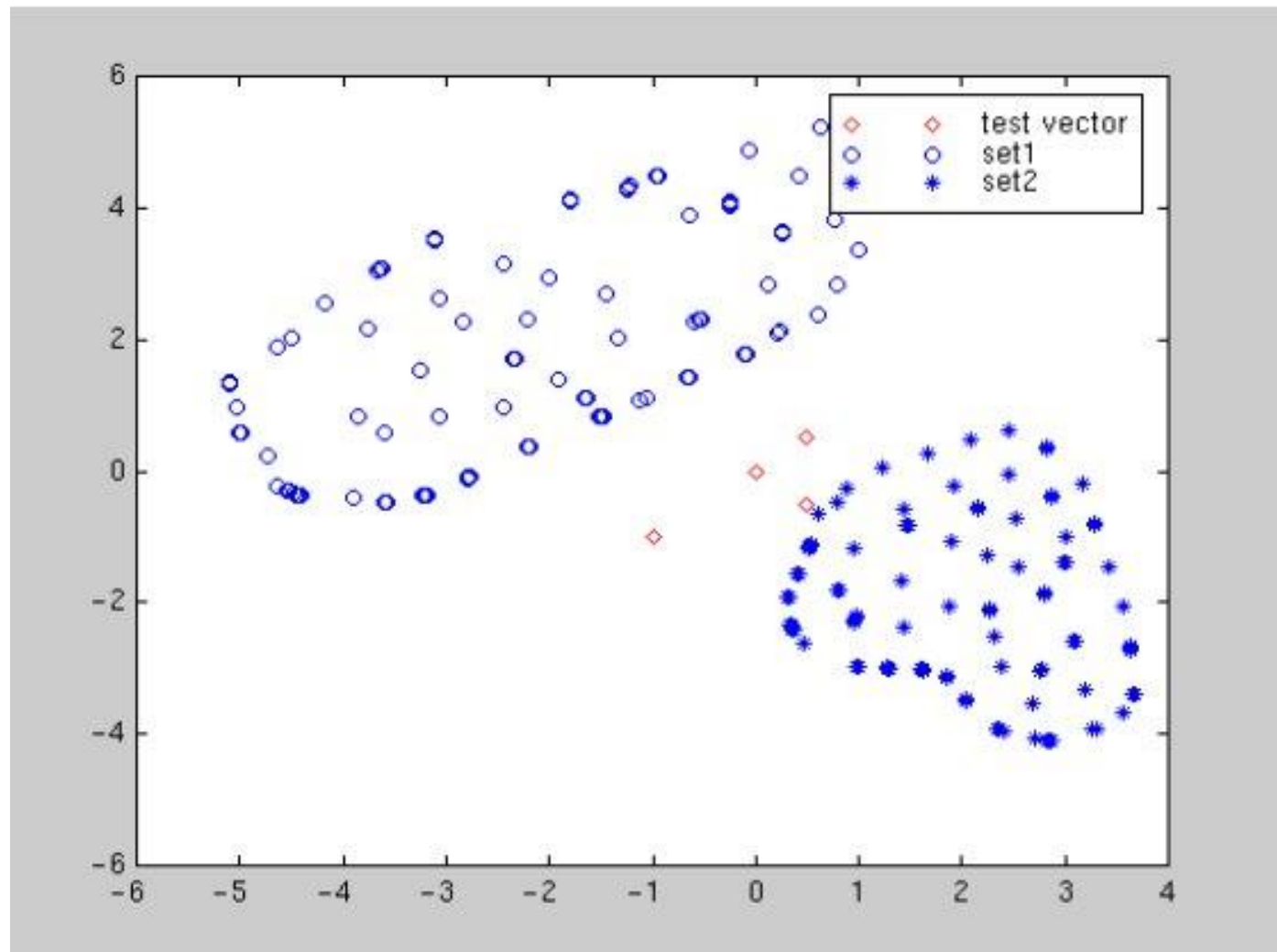
$$criterion = \frac{S_b}{S_w}$$

# Two LDA approaches

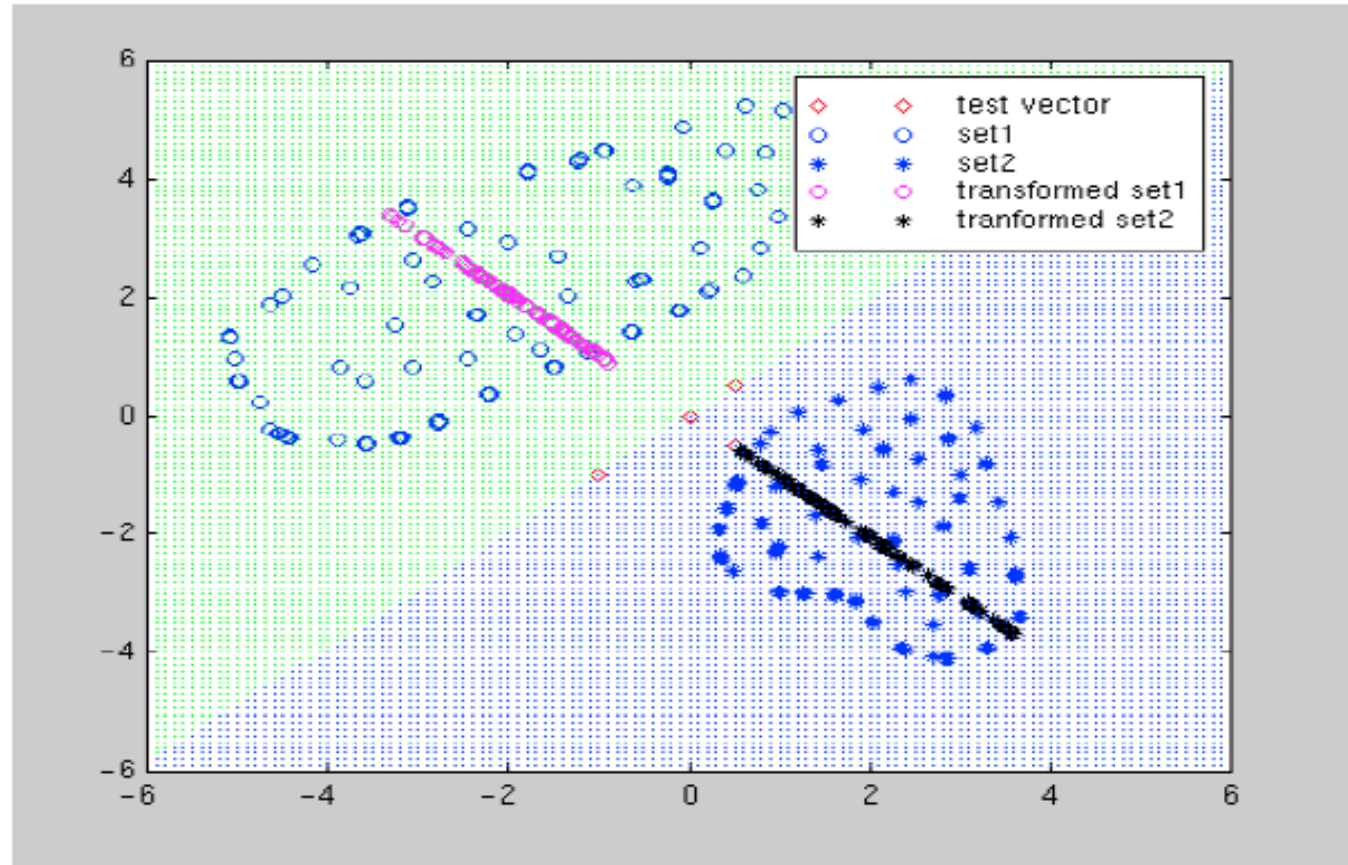
- Class-independent transformation
  - All data is transformed using the same transformation, regardless of their class
  - Maximize ratio of overall between class variance to overall within class variance  $criterion = \frac{S_b}{S_w}$
- Class-dependent transformation
  - Transform data from each class separately
  - Maximize ratio of overall between class variance to within class variance of a specific j'th class.
    - $criterion_j = \frac{S_b}{cov_j}$



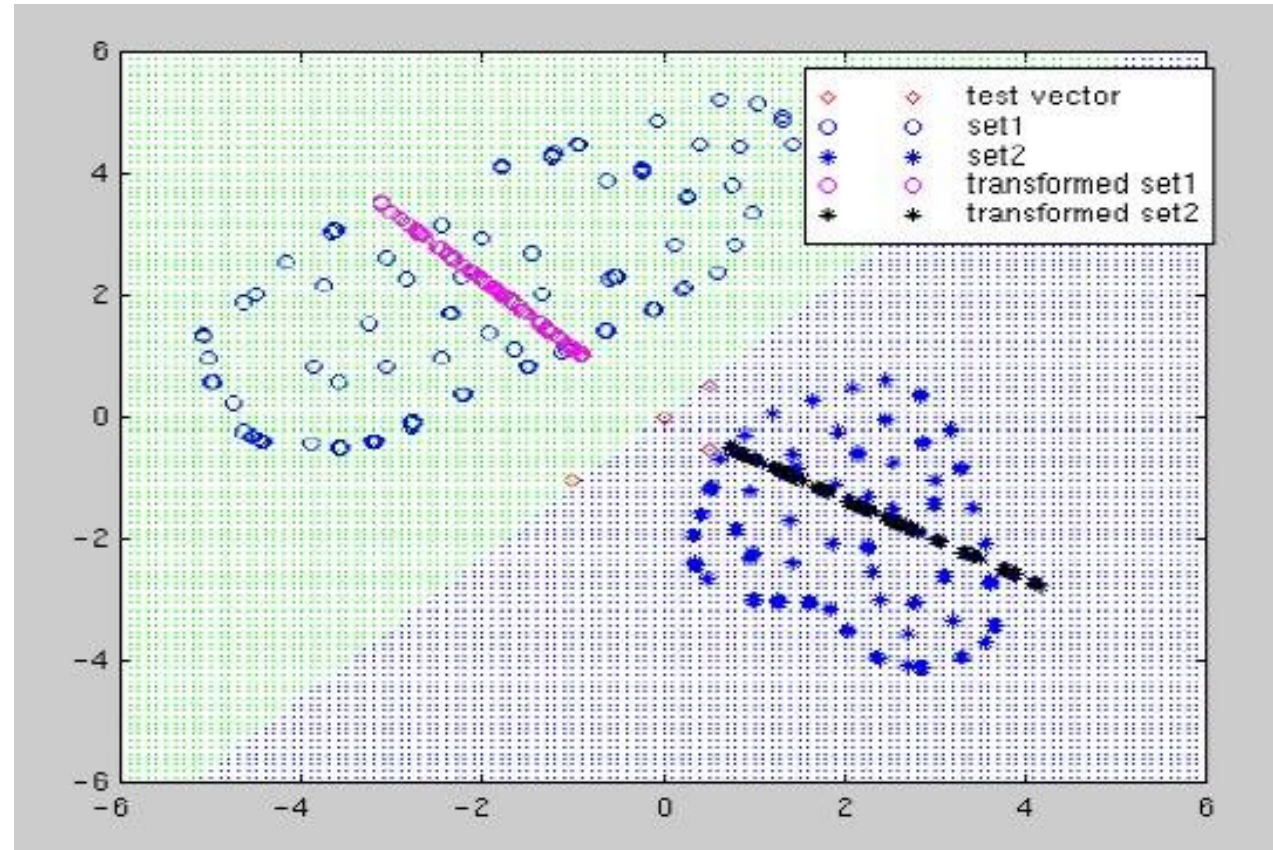
# LDA



# Class-Independent LDA



# Class-Dependent LDA



# Thoughts on LDA

- The goal in LDA is typically classification. We typically don't show the projections.
- LDA does not work well for:
  - complex problems with nonlinear patterns.
  - high dimensional data
- One solution to high dimensional data:
  - Perform PCA first, reduce to a few dimensions.
  - Then perform LDA