Feature Selection & Dimensionality Reduction

(1) Baseline: Choose some features to eliminate

(2) Transform to a new lower-dimensional feature space.

- PCA
- FLD

Principal Components Analysis (PCA)

Also: Karhunen-Loeve (KL) Transform

Use: orthonormal transformation

(1) Sample covariance matrix

$$\hat{\underline{\underline{\Sigma}}} = \frac{1}{N} \sum_{j=1}^{N} (\underline{x}_j - \underline{m})(\underline{x}_j - \underline{m})^T$$

← real & symmetric

← $D \times D$ (feature dimension)

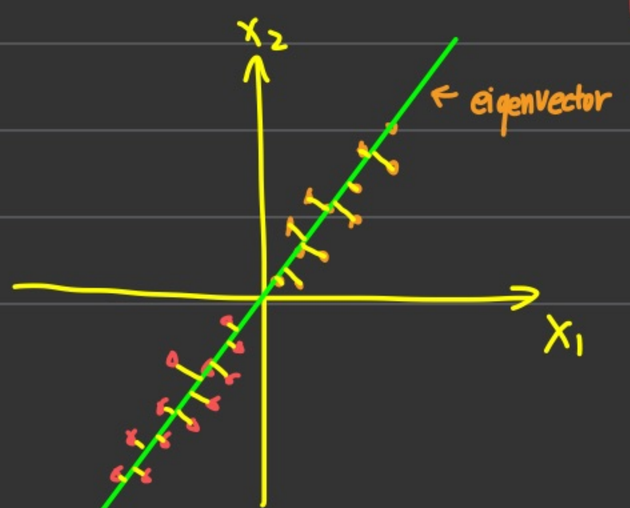
(2) Find eigenvectors $\underset{D\times 1}{\underline{e}}$ & eigenvalues λ of $\hat{\underline{\underline{\Sigma}}}$.

(3) Keep eigenvectors corresponding to D' largest eigenvalues.

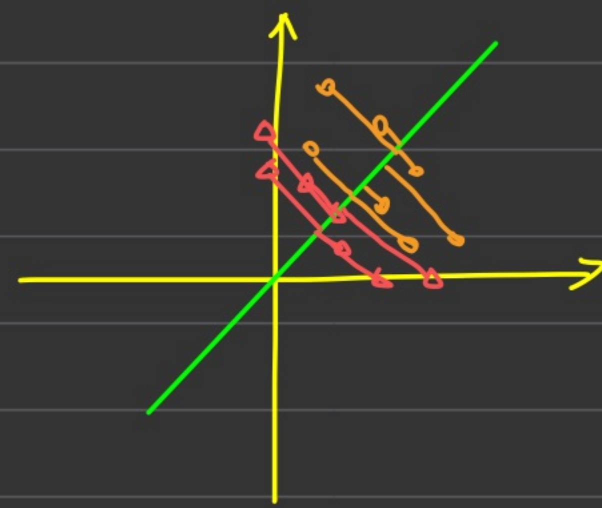
(4) Transform all data points $\underline{x}_n$ to new feature space

$$\text{new } \underset{D'\times 1}{\underline{x}_n} = \underset{D'\times D}{\underline{\underline{E}}^T} \underset{D\times 1}{\underline{x}_n}, \quad \underset{D\times D'}{\underline{\underline{E}}} = [\underbrace{\underline{e}_1 \ \underline{e}_2 \ \cdots \ \underline{e}_{D'}}]$$

each $\underline{e}_n$ is normalized to unit length.



Fisher's Linear Discriminant (FLD)



In this case, PCA works poorly.

But this one works well.

Motivation for FLD: find a direction that maximize the separability of projected data points.

$$\therefore \quad J = \frac{[\text{distance between projected class means}]^2}{[\text{variance of each projected class}]}$$

maximize it