Discriminant Functions

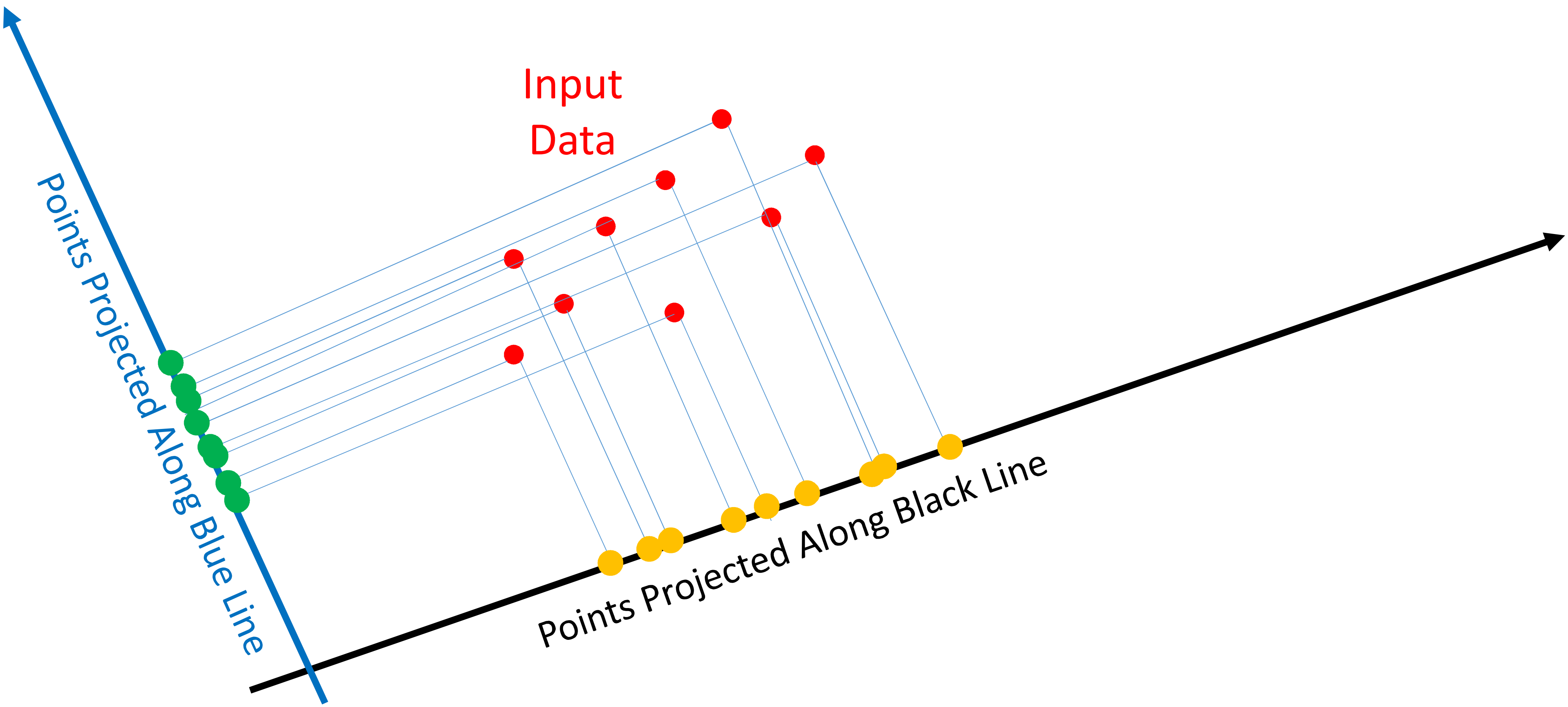
Linear Discriminant Analysis

Prithwijit Guha
Dept. of EEE, IIT Guwahati

Dimensionality Reduction

- Feature Subset Selection
- Hashing Techniques
- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Exploratory Factor Analysis (EFA)

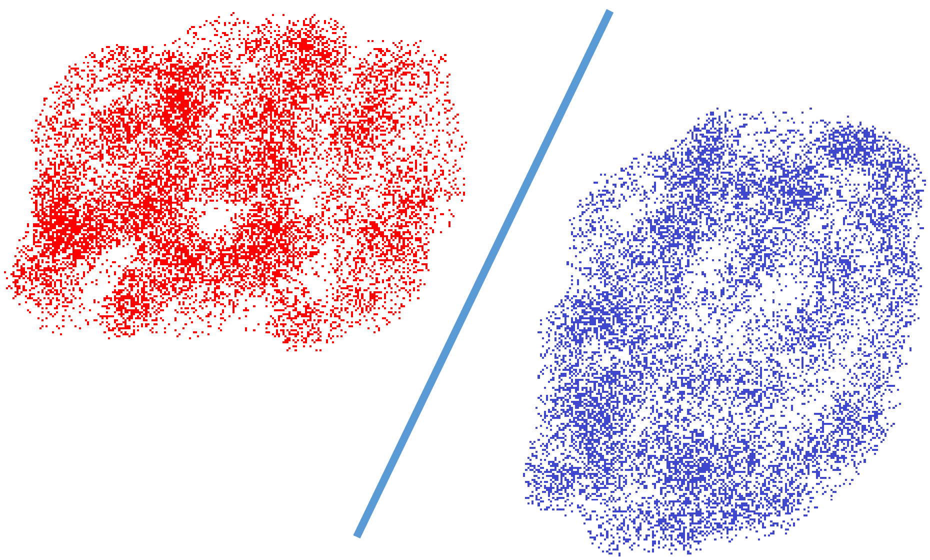
Recapitulation: Principal Components



Recapitulation: Principal Component Analysis

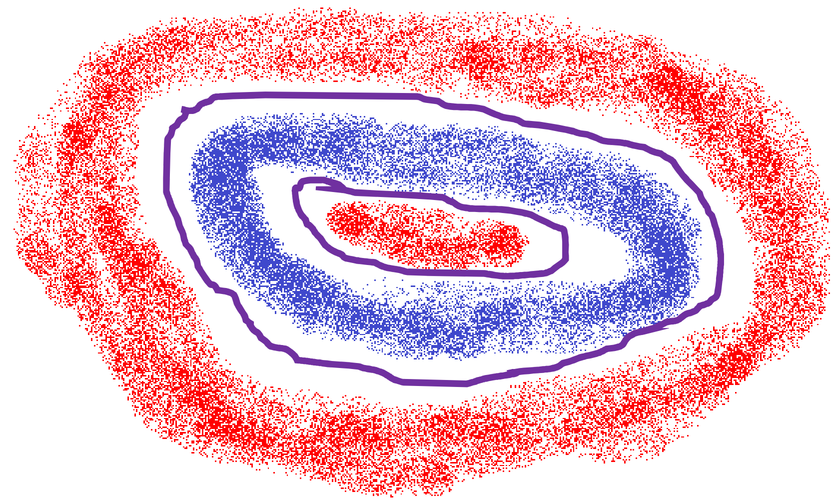
- Significance of Principal Components
- Eigen Vectors of Covariance Matrix
- Transformation Matrix from Eigen Vectors
- Provides Uncorrelated Dimensions
- Useful for Applications with
  - Highly Correlated Features
  - Features with Probably Redundant Information

Separable Classes



Linearly Separable

Not Linearly Separable



Classification: Input Data & Label

$$\boldsymbol{X_0} = \{\boldsymbol{x}_i^0 : \boldsymbol{x}_i^0 \in \mathbb{R}^D; i = 1, \dots n_0\}$$

$$\boldsymbol{X_1} = \{\boldsymbol{x}_j^1 : \boldsymbol{x}_j^1 \in \mathbb{R}^D; j = 1, \dots n_1\}$$

$$y(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{x} \in \boldsymbol{X_1} \\ 0, & \boldsymbol{x} \in \boldsymbol{X_0} \end{cases}$$

Classification: Input Distribution

$$\boldsymbol{x} \in \boldsymbol{X_0} \Rightarrow \boldsymbol{x} \sim P_0(\boldsymbol{x}; \boldsymbol{\theta}_0)$$

$$\boldsymbol{x} \in \boldsymbol{X_1} \Rightarrow \boldsymbol{x} \sim P_1(\boldsymbol{x}; \boldsymbol{\theta}_1)$$

P_0 and P_1 are the respective Probability Distributions learned from $\boldsymbol{X_0}$ and $\boldsymbol{X_1}$. The respective parameters of these Distributions are $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta_1}$.

Classification: Input Distribution

$$P(\boldsymbol{X_1}|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\boldsymbol{X_1}, \boldsymbol{\theta_1})P(\boldsymbol{X_1})}{P(\boldsymbol{x})}$$

Posterior **Likelihood** **Prior**

$$P(\boldsymbol{X_0}|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\boldsymbol{X_0}, \boldsymbol{\theta_0})P(\boldsymbol{X_0})}{P(\boldsymbol{x})}$$

Evidence

$$P(\boldsymbol{x}) = P(\boldsymbol{x}|\boldsymbol{X_0}, \boldsymbol{\theta_0})P(\boldsymbol{X_0}) + P(\boldsymbol{x}|\boldsymbol{X_1}, \boldsymbol{\theta_1})P(\boldsymbol{X_1})$$

Discriminant Functions & Decision Rule

Discriminant Function

$$y(\boldsymbol{x}) = \begin{cases} 1, & P(\boldsymbol{X_1}|\boldsymbol{x}) > P(\boldsymbol{X_0}|\boldsymbol{x}) \\ 0, & P(\boldsymbol{X_1}|\boldsymbol{x}) < P(\boldsymbol{X_0}|\boldsymbol{x}) \end{cases}$$

$$g_i(\boldsymbol{x}) = \ln\{P(\boldsymbol{X_i}|\boldsymbol{x})\}$$

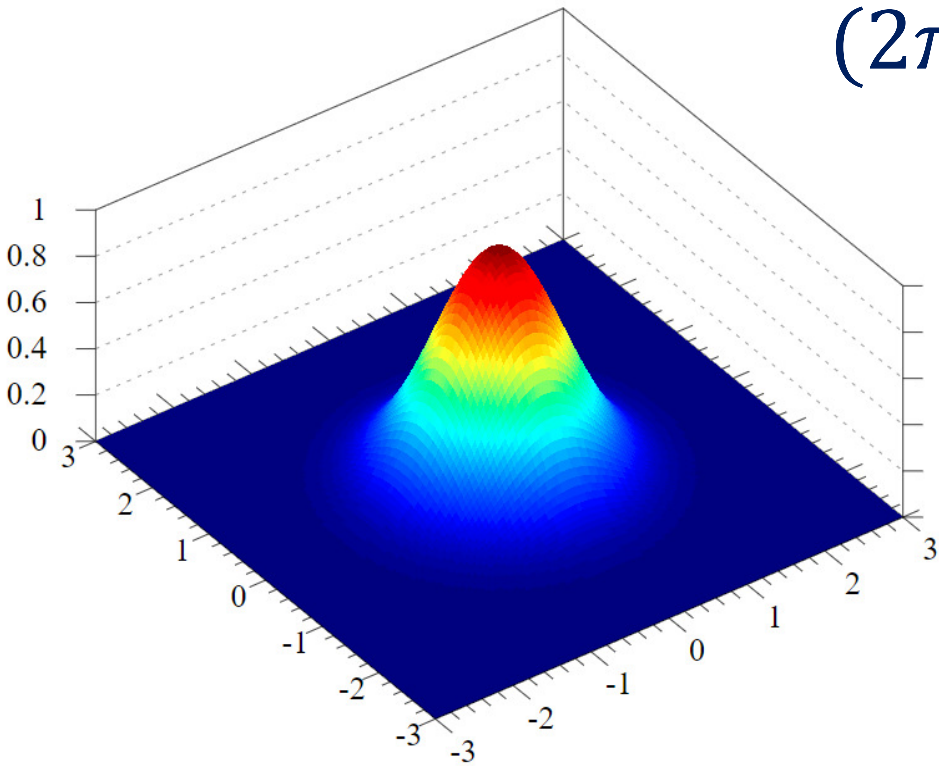
$$g(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_0(\boldsymbol{x})$$

$$\boxed{y(\boldsymbol{v}) = \begin{cases} 1, & g(\boldsymbol{v}) = g_1(\boldsymbol{v}) - g_0(\boldsymbol{v}) > 0 \\ 0, & g(\boldsymbol{v}) = g_1(\boldsymbol{v}) - g_0(\boldsymbol{v}) < 0 \end{cases}}$$

Classification Decision Rule (unseen data $\boldsymbol{v}$)

Discriminant Functions: Gaussian Distribution

$$P(\boldsymbol{x};\boldsymbol{\theta}=[\boldsymbol{\mu},\boldsymbol{C}])=\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$



Case-1: $\boldsymbol{C}_1=\boldsymbol{C}_2=\sigma^2\boldsymbol{I}$

Case-2: $\boldsymbol{C}_1=\boldsymbol{C}_2=\boldsymbol{C}$

Case-3: $\boldsymbol{C}_1\neq\boldsymbol{C}_2$

Discriminant Function: Gaussian Distribution

$$g_i(\boldsymbol{x}) = ln\{P(\boldsymbol{X_i}|\boldsymbol{x})\} = ln\left\{\frac{P(\boldsymbol{x}|\boldsymbol{X_i},\boldsymbol{\theta_i})P(\boldsymbol{X_i})}{P(\boldsymbol{x})}\right\}$$

$$= ln\{P(\boldsymbol{x}|\boldsymbol{X}_i,\boldsymbol{\theta}_i)\} + ln\{P(\boldsymbol{X}_i)\} - ln\{P(\boldsymbol{x})\}$$

$$\boxed{\begin{aligned} g_i(\boldsymbol{x}) = -\frac{n}{2}ln\{2\pi\} - \frac{1}{2}ln\{|\boldsymbol{C}_i|\} - \frac{1}{2}(x-\boldsymbol{\mu}_i)^T\boldsymbol{C}_i^{-1}(x-\boldsymbol{\mu}_i) \\ + ln\{P(\boldsymbol{X_i})\} - ln\{P(\boldsymbol{x})\} \end{aligned}}$$

Discriminant Function: Gaussian Distribution

$$g_i(\boldsymbol{x}) = -\frac{n}{2} ln\{2\pi\} - \frac{1}{2} ln\{|\boldsymbol{C}_i|\} - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^T \boldsymbol{C}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) + ln\{P(\boldsymbol{X_i})\} - ln\{P(\boldsymbol{x})\}$$

$$\boxed{g(\boldsymbol{x}) = -\frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x} - \boldsymbol{\mu}_1)^T \boldsymbol{C}_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) - (\boldsymbol{x} - \boldsymbol{\mu}_0)^T \boldsymbol{C}_0^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_0)\}}$$

Discriminant Function: $\boldsymbol{C_1 = C_0 = \sigma^2 I}$

$$g(\boldsymbol{x}) = -\frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \boldsymbol{C}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T \boldsymbol{C}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= -\frac{1}{2} ln\left\{\frac{\sigma^2}{\sigma^2}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T(\sigma^{-2}\boldsymbol{I})\,(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T(\sigma^{-2}\boldsymbol{I})(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= 0 + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2\sigma^2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\,(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T\,(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= -\frac{1}{2\sigma^2}\{(\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_1^T\,\boldsymbol{x} + \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1) - (\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_0^T\,\boldsymbol{x} + \boldsymbol{\mu}_0^T\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

$$= -\frac{1}{2\sigma^2}\{-2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\boldsymbol{x} + (\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

Discriminant Function: $\boldsymbol{C_1} = \boldsymbol{C_0} = \sigma^2 \boldsymbol{I}$

$$g(\boldsymbol{x}) = -\frac{1}{2\sigma^2}\{-2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{x} + (\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

$$g(\boldsymbol{x}) = \frac{1}{\sigma^2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{x} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2\sigma^2}(\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\mu}_0)$$

$$g(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$$

$$\boldsymbol{a} = \frac{1}{\sigma^2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

$$b = ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2\sigma^2}(\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\mu}_0)$$

Discriminant Function: $\boldsymbol{C_1 = C_0 = C}$

$$g(\boldsymbol{x}) = -\frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{C}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{C}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= -\frac{1}{2} ln\left\{\frac{|\boldsymbol{C}|}{|\boldsymbol{C}|}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= -\frac{1}{2}\{(\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1) - (\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0)\}$$
$$+ ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

$$= -\frac{1}{2}\{-2\boldsymbol{x}^T\boldsymbol{C}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) + (\boldsymbol{\mu}_1^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

Discriminant Function: $\boldsymbol{C_1} = \boldsymbol{C_0} = \boldsymbol{C}$

$$g(\boldsymbol{x}) = -\frac{1}{2}\{-2\boldsymbol{x}^T\boldsymbol{C}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) + (\boldsymbol{\mu}_1^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$

$$g(\boldsymbol{x}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{C}^{-1}\boldsymbol{x} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0)$$

$$g(\boldsymbol{x}) = \boldsymbol{p}^T\boldsymbol{x} + q$$

$$\boldsymbol{p} = \boldsymbol{C}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

$$q = ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}^{-1}\boldsymbol{\mu}_0)$$

Discriminant Function: $\boldsymbol{C_1} \neq \boldsymbol{C_0}$

$$g(\boldsymbol{x}) = -\frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{C}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{C}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\}$$

$$= -\frac{1}{2}\{(\boldsymbol{x}^T\boldsymbol{C}_1^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1) - (\boldsymbol{x}^T\boldsymbol{C}_0^{-1}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0)\}$$
$$+ ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\}$$

$$= -\frac{1}{2}\{\boldsymbol{x}^T(\boldsymbol{C}_1^{-1} - \boldsymbol{C}_0^{-1})\boldsymbol{x} - 2\boldsymbol{x}^T(\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0) + (\boldsymbol{\mu}_1^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0)\} + ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\}$$
$$- \frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\}$$

Discriminant Function: $\boldsymbol{C_1} \neq \boldsymbol{C_0}$

$$g(\boldsymbol{x}) = -\frac{1}{2}\{\boldsymbol{x}^T(\boldsymbol{C}_1^{-1} - \boldsymbol{C}_0^{-1})\boldsymbol{x} - 2\boldsymbol{x}^T(\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0) + (\boldsymbol{\mu}_1^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0)\}$$
$$+ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\})$$

$$g(\boldsymbol{x}) = \left[ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0) - \frac{1}{2}ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\}\right] + \boldsymbol{x}^T(\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0)$$
$$-\frac{1}{2}\boldsymbol{x}^T(\boldsymbol{C}_1^{-1} - \boldsymbol{C}_0^{-1})\boldsymbol{x}$$

Discriminant Function: $\boldsymbol{C_1} \neq \boldsymbol{C_0}$

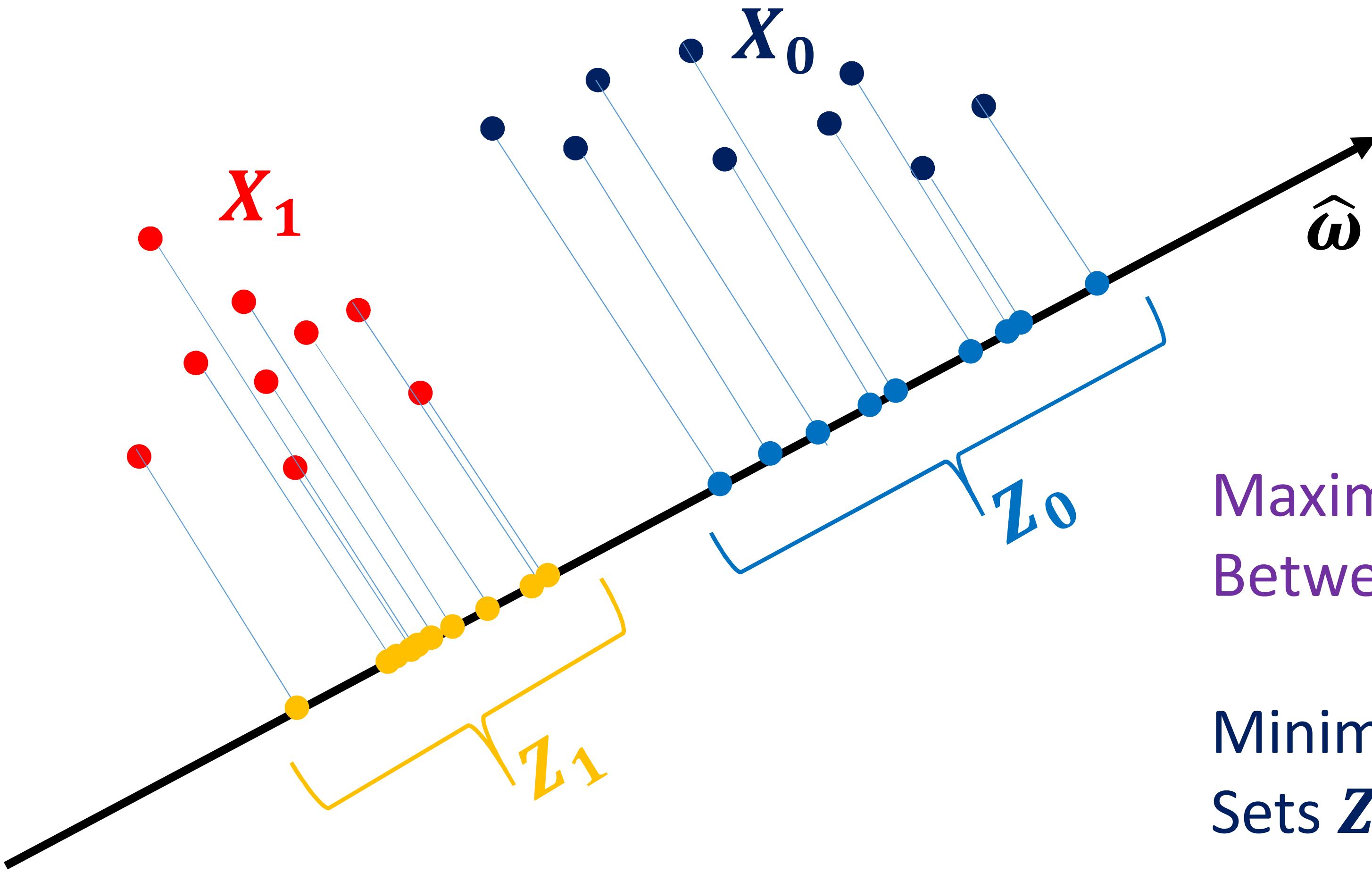
$$g(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + c$$

$$\boldsymbol{A} = \boldsymbol{C}_1^{-1} - \boldsymbol{C}_0^{-1}$$

$$\boldsymbol{b} = \boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0$$

$$c = \left[ln\left\{\frac{P(\boldsymbol{X_1})}{P(\boldsymbol{X_0})}\right\} - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{C}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0) - \frac{1}{2} ln\left\{\frac{|\boldsymbol{C}_1|}{|\boldsymbol{C}_0|}\right\}\right]$$

Linear Discriminant Analysis (LDA)



Maximize Distance Between Sets $\boldsymbol{Z_0}$ and $\boldsymbol{Z_1}$

Minimize Variances of Sets $\boldsymbol{Z_0}$ and $\boldsymbol{Z_1}$

Recapitulation: Evaluating Mean and Variance

$$\boldsymbol{X} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathbb{R}^D; i = 1, \ldots n\} \qquad \boldsymbol{p}^T\boldsymbol{p} = 1$$

$$\boldsymbol{Z} = \{z_i : z_i = \boldsymbol{p}^T\boldsymbol{x}_i; \forall \boldsymbol{x}_i \in \boldsymbol{X}; z_i \in \mathbb{R}^1\}$$

$$\gamma = Mean(\boldsymbol{Z}) = \frac{1}{n}\sum_{i}^{n} z_i \qquad \boldsymbol{\mu} = Mean(\boldsymbol{X}) = \frac{1}{n}\sum_{i}^{n} \boldsymbol{x}_i$$

$$\gamma = \frac{1}{n}\sum_{i}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{p}^T\boldsymbol{x}_i = p^T\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\right) \Rightarrow \gamma = \boldsymbol{p}^T\boldsymbol{\mu}$$

Recapitulation: Evaluating Mean and Variance

$$\boldsymbol{C} = Cov(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

$$v(\boldsymbol{p}) = Var(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^{n}(z_i - \gamma)^2 = \frac{1}{n}\sum_{i=1}^{n}z_i^2 - \gamma^2$$

$$v(\boldsymbol{p}) = \frac{1}{n}\sum_{i=1}^{n}z_i^2 - \gamma^2 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{p}^T\boldsymbol{x}_i)(\boldsymbol{x}_i^T\boldsymbol{p}) - (\boldsymbol{p}^T\boldsymbol{\mu})(\boldsymbol{\mu}^T\boldsymbol{p})$$

$$v(\boldsymbol{p}) = \boldsymbol{p}^T\left\{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i)(\boldsymbol{x}_i^T) - (\boldsymbol{\mu})(\boldsymbol{\mu}^T)\right\}\boldsymbol{p} \Rightarrow v(\boldsymbol{p}) = \boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p}$$

Definitions

$$\boldsymbol{X_0} = \{\boldsymbol{x}_i^0 : \boldsymbol{x}_i^0 \in \mathbb{R}^D; i = 1, \dots n_0\}$$

$$\boldsymbol{X_1} = \{\boldsymbol{x}_j^1 : \boldsymbol{x}_j^1 \in \mathbb{R}^D; j = 1, \dots n_1\}$$

$$y(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{x} \in \boldsymbol{X_1} \\ 0, & \boldsymbol{x} \in \boldsymbol{X_0} \end{cases}$$

$$\boldsymbol{Z_0} = \{z_i^0 : z_i^0 = \boldsymbol{\omega}^T \boldsymbol{x}_i^0; \boldsymbol{x}_i^0 \in \boldsymbol{X_0}\}$$

$$\boldsymbol{Z_1} = \{z_j^1 : z_j^1 = \boldsymbol{\omega}^T \boldsymbol{x}_j^1; \boldsymbol{x}_j^1 \in \boldsymbol{X_1}\}$$

$$\boldsymbol{\omega}^T \boldsymbol{\omega} = 1$$

$$\boldsymbol{\mu}_0 = Mean(\boldsymbol{X_0}); \boldsymbol{C}_0 = Cov(\boldsymbol{X_0}) \qquad m_0 = Mean(\boldsymbol{Z_0}); v_0 = Var(\boldsymbol{Z_0})$$

$$\boldsymbol{\mu}_1 = Mean(\boldsymbol{X_1}); \boldsymbol{C}_1 = Cov(\boldsymbol{X_1}) \qquad m_1 = Mean(\boldsymbol{Z_1}); v_1 = Var(\boldsymbol{Z_1})$$

Problem Formulation: Step #1

Maximize Distance between Sets $\boldsymbol{Z_0}$ and $\boldsymbol{Z_1}$

$$Maximize: (m_1 - m_0)^2 = \boldsymbol{\omega}^T \boldsymbol{S_B} \boldsymbol{\omega}$$

$$(m_1 - m_0)^2 = (\boldsymbol{\omega}^T \boldsymbol{\mu}_1 - \boldsymbol{\omega}^T \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1^T \boldsymbol{\omega} - \boldsymbol{\mu}_0^T \boldsymbol{\omega})$$

$$= \boldsymbol{\omega}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\omega}$$

Between-Class ($\boldsymbol{X_0}$ and $\boldsymbol{X_1}$) Scatter Matrix

$$\boldsymbol{S_B} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T$$

Problem Formulation: Step #2

Minimize Net Variances of Sets $\boldsymbol{Z_0}$ and $\boldsymbol{Z_1}$

$$Minimize: v_1 + v_0 = \boldsymbol{\omega}^T \boldsymbol{S}_W \boldsymbol{\omega}$$

$$v_1 + v_0 = \boldsymbol{\omega}^T \boldsymbol{C}_1 \boldsymbol{\omega} + \boldsymbol{\omega}^T \boldsymbol{C}_0 \boldsymbol{\omega} = \boldsymbol{\omega}^T (\boldsymbol{C}_1 + \boldsymbol{C}_0) \boldsymbol{\omega}$$

Net Within-Class ($\boldsymbol{X_0}$ and $\boldsymbol{X_1}$) Scatter Matrix

$$\boldsymbol{S}_W = \boldsymbol{C}_1 + \boldsymbol{C}_0$$

LDA: Objective

$$Maximize: \boldsymbol{\omega}^T \boldsymbol{S_B} \boldsymbol{\omega}$$

$$Minimize: \boldsymbol{\omega}^T \boldsymbol{S_W} \boldsymbol{\omega}$$

$$Maximize: \frac{\boldsymbol{\omega}^T \boldsymbol{S_B} \boldsymbol{\omega}}{\boldsymbol{\omega}^T \boldsymbol{S_W} \boldsymbol{\omega}} \quad w.r.t. \boldsymbol{\omega}$$

$$g(\boldsymbol{\omega}; \lambda) = -\boldsymbol{\omega}^T \boldsymbol{S_B} \boldsymbol{\omega} + \lambda(\boldsymbol{\omega}^T \boldsymbol{S_W} \boldsymbol{\omega} - 1)$$

Recapitulation: KKT Conditions

Minimize: $f(\boldsymbol{x})$ Subject to: $h_i(\boldsymbol{x}) = 0; i = 1, \ldots n$

Lagrangian: $\mathrm{g}(\boldsymbol{x}; \alpha_1, \ldots \alpha_n) = f(\boldsymbol{x}) + \sum_{i=1}^{n} \alpha_i h_i(\boldsymbol{x})$

KKT Conditions: $\nabla_x \mathrm{g}(\boldsymbol{x}; \alpha_1, \ldots \alpha_n) = \nabla_x f(\boldsymbol{x}) + \sum_{i=1}^{n} \alpha_i \nabla_x h_i(\boldsymbol{x}) = 0$

Maximize: $f(\boldsymbol{x})$ Subject to: $h_i(\boldsymbol{x}) = 0; i = 1, \ldots n$

Lagrangian: $\mathrm{g}(\boldsymbol{x}; \alpha_1, \ldots \alpha_n) = -f(\boldsymbol{x}) + \sum_{i=1}^{n} \alpha_i h_i(\boldsymbol{x})$

KKT Conditions: $\nabla_x \mathrm{g}(\boldsymbol{x}; \alpha_1, \ldots \alpha_n) = -\nabla_x f(\boldsymbol{x}) + \sum_{i=1}^{n} \alpha_i \nabla_x h_i(\boldsymbol{x}) = 0$

LDA: Solution

$$g(\boldsymbol{\omega};\lambda) = -\boldsymbol{\omega}^T \boldsymbol{S_B}\boldsymbol{\omega} + \lambda(\boldsymbol{\omega}^T \boldsymbol{S_W}\boldsymbol{\omega} - 1)$$

Application of KKT Conditions

$$\nabla_{\boldsymbol{\omega}} g(\boldsymbol{\omega};\lambda) = -2\boldsymbol{S_B}\boldsymbol{\omega} + \lambda(2\boldsymbol{S_W}\boldsymbol{\omega}) = 0$$

$$\Downarrow$$

$$2\boldsymbol{S_B}\boldsymbol{\omega} = 2\lambda\boldsymbol{S_W}\boldsymbol{\omega}$$

$$\Downarrow$$

$$\boldsymbol{S_W^{-1}}\boldsymbol{S_B}\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$$

Illustration: $A = \boldsymbol{u}\boldsymbol{u}^T$ is Rank-1

$$\boldsymbol{u} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \qquad A = \boldsymbol{u}\boldsymbol{u}^T = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} -1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \\ -2 & 2 & 4 \end{bmatrix}$$

$$eigVec(A) = \begin{bmatrix} -0.41 & 0.71 & -0.58 \\ 0.41 & 0.71 & 0.58 \\ 0.82 & 0 & -0.58 \end{bmatrix}$$

$$eigVal(A) = \begin{bmatrix} 6 & 0 & 0 \end{bmatrix}$$

Illustration: $\mathrm{M} = B * (\boldsymbol{u}\boldsymbol{u}^T)$ is Also Rank-1

$$B = \begin{bmatrix} 3 & 2 & -1 \\ 5 & -7 & 2 \\ 4 & 1 & -3 \end{bmatrix} \longrightarrow eigVal(B) = [3.51 \quad -2.48 \quad -8.03]$$

$$A = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \\ -2 & 2 & 4 \end{bmatrix}$$

$$eigVec(M) = \begin{bmatrix} -0.24 & -0.76 & -0.66 \\ -0.64 & 0.35 & -0.75 \\ -0.73 & -0.55 & 0.04 \end{bmatrix}$$

$$M = B * A$$

$$M = \begin{bmatrix} 3 & -3 & -6 \\ 8 & -8 & -16 \\ 9 & -9 & -18 \end{bmatrix} \longrightarrow eigVal(M) = [-23 \quad 0 \quad 0]$$

LDA: Solution

$$\boldsymbol{S}_W^{-1}\boldsymbol{S}_B\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$$

We are only Interested in the Unit Direction $\hat{\boldsymbol{\omega}}$

$$\lambda\boldsymbol{\omega} = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\omega} = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\omega}\}$$

$$\left(\frac{\lambda}{\alpha}\right)\boldsymbol{\omega} = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0);\ \alpha = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\omega}$$

$$\Downarrow$$

$$\boldsymbol{\omega} \propto \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

Example: Input Data

X_1	**1**	**2**	**3**	**5**	**4**	**6**	**8**
	1	**2**	**3**	**4**	**5**	**6**	**7**

X_0	**-2**	**-1**	**1**	**3**	**4**	**2**	**5**
	3	4	5	6	7	8	9

Example: Compute Means

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 4.1429 \\ 4.0000 \end{pmatrix} \qquad \boldsymbol{\mu}_0 = \begin{pmatrix} 1.7143 \\ 6.0000 \end{pmatrix}$$

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = \begin{pmatrix} 2.4286 \\ -2.000 \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} 2.9286 \\ 5.0000 \end{pmatrix}$$

Example: Compute Covariances

$$\boldsymbol{C}_1 = \begin{pmatrix} 4.9796 & 4.2857 \\ 4.2857 & 4.0000 \end{pmatrix} \qquad \boldsymbol{C}_0 = \begin{pmatrix} 5.6327 & 4.2857 \\ 4.2857 & 4.0000 \end{pmatrix}$$

$$\boldsymbol{S}_W = \boldsymbol{C}_1 + \boldsymbol{C}_0 = \begin{pmatrix} 10.6122 & 8.5714 \\ 8.5714 & 8.0000 \end{pmatrix}$$

$$\boldsymbol{C} = \begin{pmatrix} 6.7806 & 3.0714 \\ 3.0714 & 5.0000 \end{pmatrix}$$

Example: LDA

$$\boldsymbol{u} = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \begin{pmatrix} 3.2000 \\ -3.6786 \end{pmatrix}$$

$$\widehat{\boldsymbol{\omega}} = \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_2} = \begin{pmatrix} 0.6563 \\ -0.7545 \end{pmatrix}$$

Comparison: LDA & PCA Directions

Best Principal Component

$$C \rightarrow \begin{pmatrix} \boxed{\begin{matrix} -\mathbf{0.7995} \\ -\mathbf{0.6007} \end{matrix}} & \begin{matrix} \mathbf{0.6007} \\ -\mathbf{0.7995} \end{matrix} \\ \multicolumn{2}{|c|}{\mathbf{9.0882} \qquad \mathbf{2.6924}} \end{pmatrix}$$

Eigen Values

$$\left\langle \begin{pmatrix} -0.7995 \\ -0.6007 \end{pmatrix}, \begin{pmatrix} 0.6563 \\ -0.7545 \end{pmatrix} \right\rangle = -0.0715$$

For this Particular Problem, LDA Direction and Best Principal Component are Almost Orthogonal

Summary

- Discriminant Functions
- LDA: Supervised Algorithm
- Use of Label Information
- Maximize Distance with Compact Groups

Thank You