

Linear Discriminant Analysis

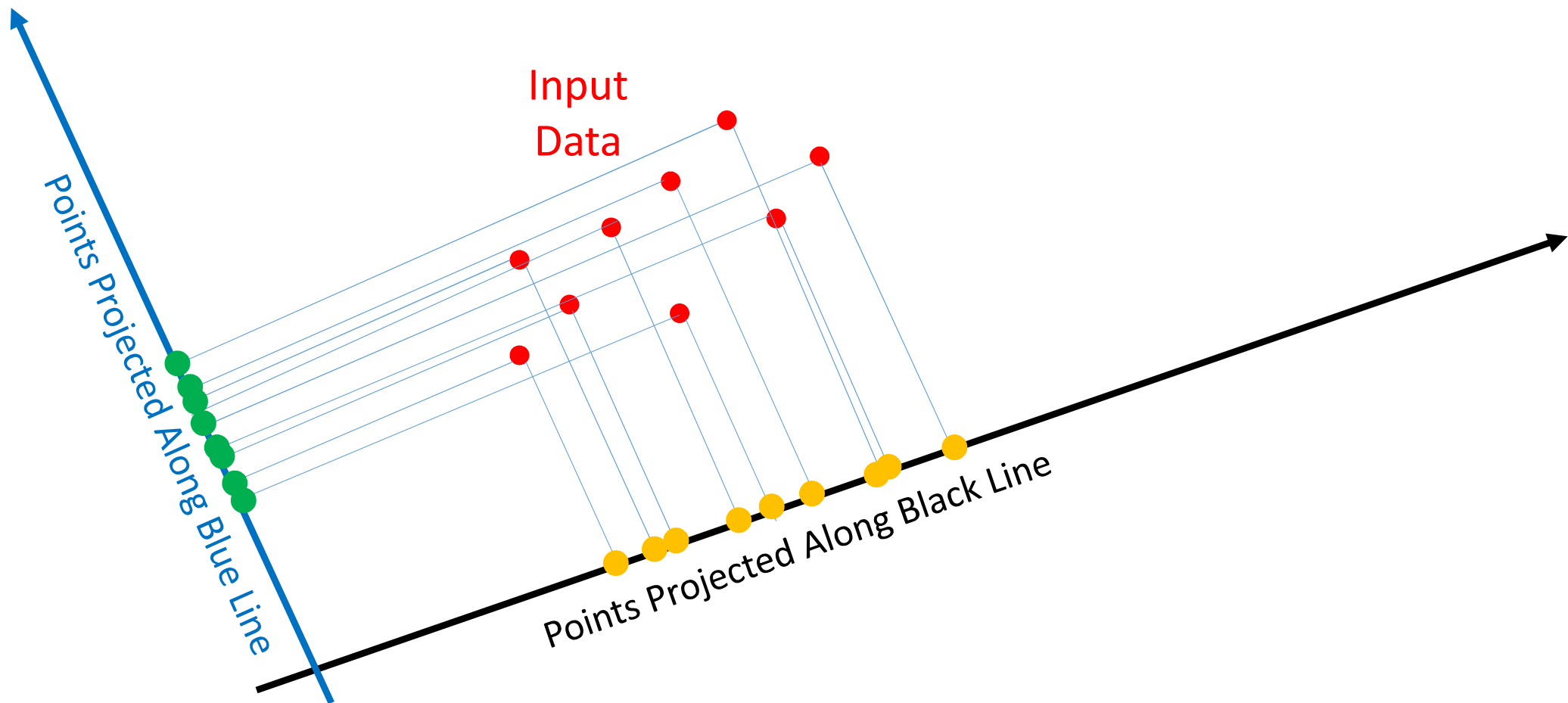


Prithwijit Guha
Dept. of EEE, IIT Guwahati

Dimensionality Reduction

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

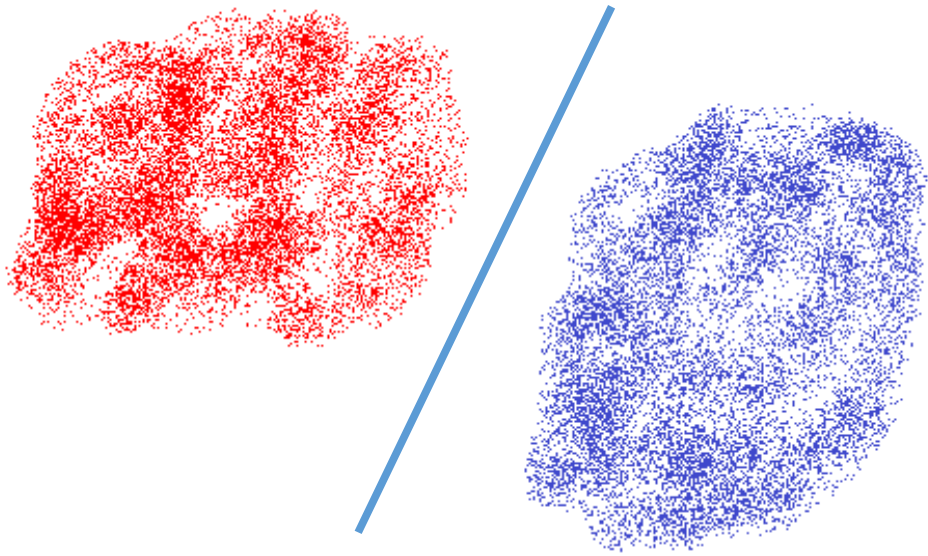
Recapitulation: Principal Components



Recapitulation: Principal Component Analysis

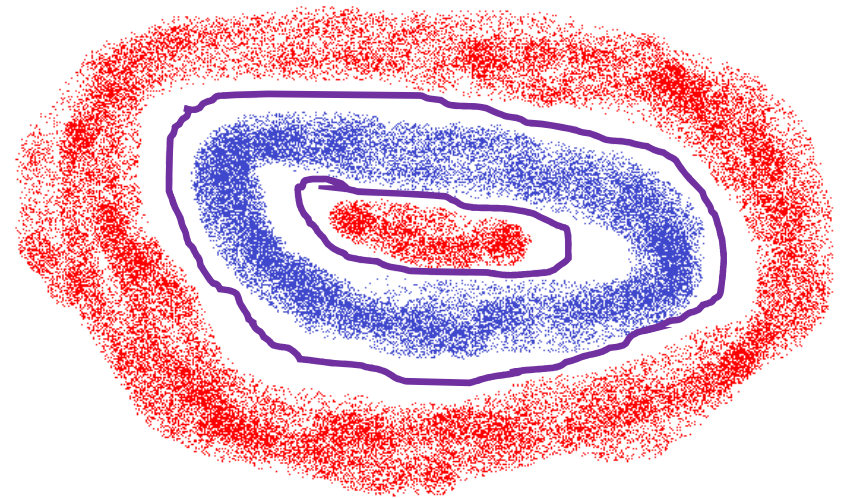
- Significance of Principal Components
- Eigen Vectors of Covariance Matrix
- Transformation Matrix from Eigen Vectors
- Provides Uncorrelated Dimensions
- Useful for Applications with
 - Highly Correlated Features
 - Features with Probably Redundant Information

Separable Classes



Linearly Separable

Not Linearly
Separable



Classification: Input Data & Label

$$\mathbf{X}_0 = \{\mathbf{x}_i^0 : \mathbf{x}_i^0 \in \mathbb{R}^D; i = 1, \dots, n_0\}$$

$$\mathbf{X}_1 = \{\mathbf{x}_j^1 : \mathbf{x}_j^1 \in \mathbb{R}^D; j = 1, \dots, n_1\}$$

$$y(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathbf{X}_1 \\ 0, & \mathbf{x} \in \mathbf{X}_0 \end{cases}$$

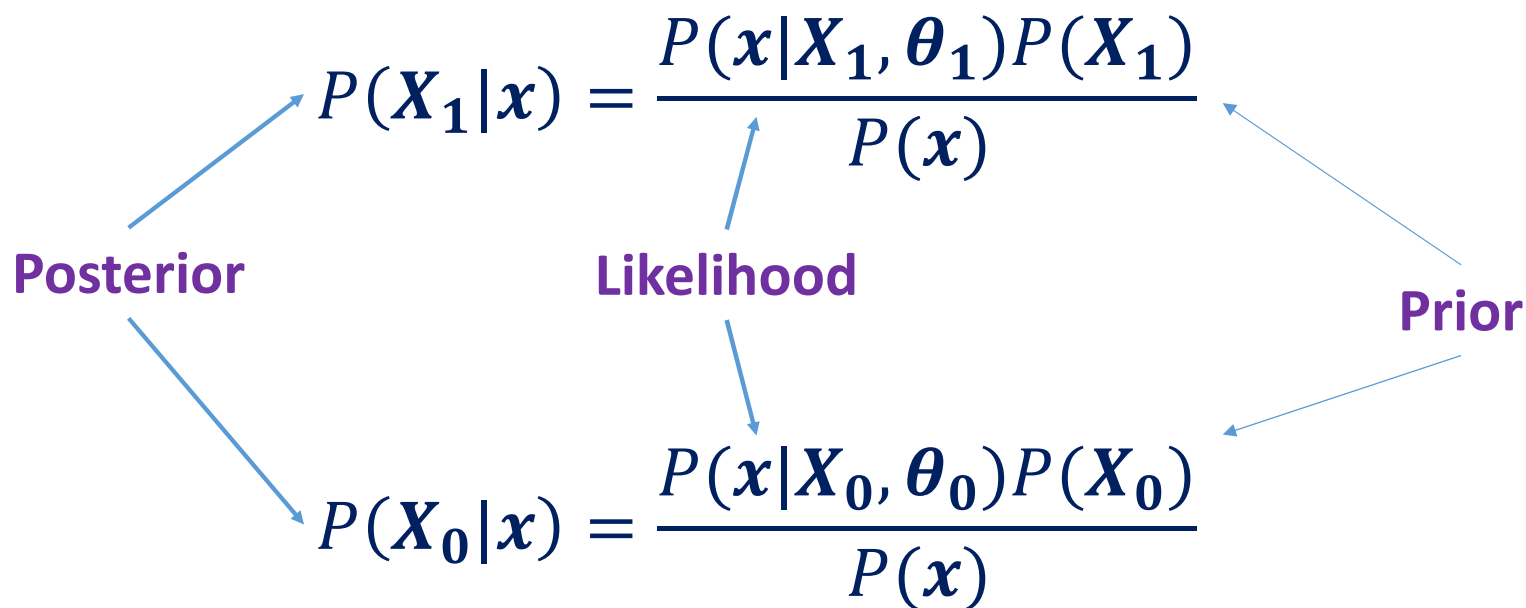
Classification: Input Distribution

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

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

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




Evidence

$$P(\mathbf{x}) = P(\mathbf{x}|X_0, \boldsymbol{\theta}_0)P(X_0) + P(\mathbf{x}|X_1, \boldsymbol{\theta}_1)P(X_1)$$

Discriminant Functions & Decision Rule

$$y(\mathbf{x}) = \begin{cases} 1, & P(\mathbf{X}_1|\mathbf{x}) > P(\mathbf{X}_0|\mathbf{x}) \\ 0, & P(\mathbf{X}_1|\mathbf{x}) < P(\mathbf{X}_0|\mathbf{x}) \end{cases}$$

Discriminant Function
 $g_i(\mathbf{x}) = \ln\{P(\mathbf{X}_i|\mathbf{x})\}$
 $g(\mathbf{x}) = g_1(\mathbf{x}) - g_0(\mathbf{x})$

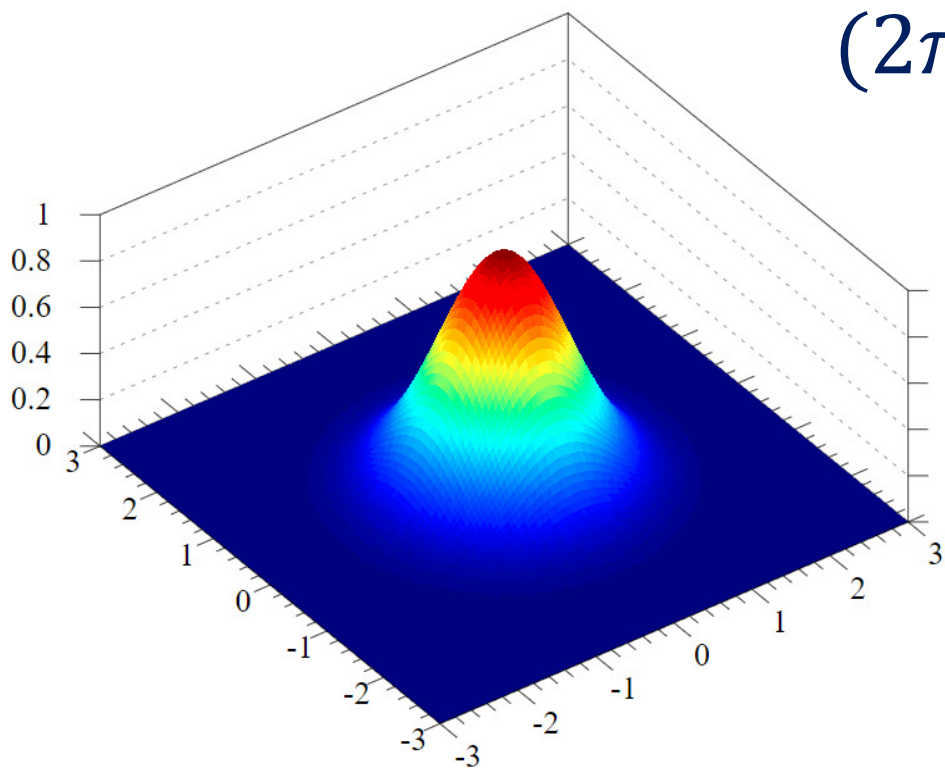


$$y(\mathbf{v}) = \begin{cases} 1, & g(\mathbf{v}) = g_1(\mathbf{v}) - g_0(\mathbf{v}) > 0 \\ 0, & g(\mathbf{v}) = g_1(\mathbf{v}) - g_0(\mathbf{v}) < 0 \end{cases}$$

Classification Decision Rule (unseen data \mathbf{v})

Discriminant Functions: Gaussian Distribution

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



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

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

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

Discriminant Function: Gaussian Distribution

$$g_i(\mathbf{x}) = \ln\{P(\mathbf{X}_i|\mathbf{x})\} = \ln\left\{\frac{P(\mathbf{x}|\mathbf{X}_i, \boldsymbol{\theta}_i)P(\mathbf{X}_i)}{P(\mathbf{x})}\right\}$$
$$= \ln\{P(\mathbf{x}|\mathbf{X}_i, \boldsymbol{\theta}_i)\} + \ln\{P(\mathbf{X}_i)\} - \ln\{P(\mathbf{x})\}$$

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

Discriminant Function: Gaussian Distribution

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

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

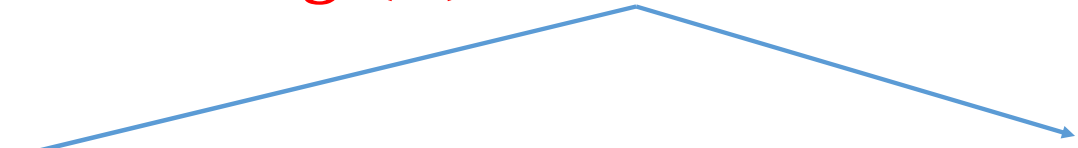
Discriminant Function: $\mathbf{C}_1 = \mathbf{C}_0 = \sigma^2 \mathbf{I}$

$$\begin{aligned} g(\mathbf{x}) &= -\frac{1}{2} \ln \left\{ \frac{|\mathbf{C}_1|}{|\mathbf{C}_0|} \right\} + \ln \left\{ \frac{P(\mathbf{X}_1)}{P(\mathbf{X}_0)} \right\} - \frac{1}{2} \{ (\mathbf{x} - \boldsymbol{\mu}_1)^T \mathbf{C}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_0)^T \mathbf{C}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \} \\ &= -\frac{1}{2} \ln \left\{ \frac{\sigma^2}{\sigma^2} \right\} + \ln \left\{ \frac{P(\mathbf{X}_1)}{P(\mathbf{X}_0)} \right\} - \frac{1}{2} \{ (\mathbf{x} - \boldsymbol{\mu}_1)^T (\sigma^{-2} \mathbf{I}) (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_0)^T (\sigma^{-2} \mathbf{I}) (\mathbf{x} - \boldsymbol{\mu}_0) \} \\ &= 0 + \ln \left\{ \frac{P(\mathbf{X}_1)}{P(\mathbf{X}_0)} \right\} - \frac{1}{2\sigma^2} \{ (\mathbf{x} - \boldsymbol{\mu}_1)^T (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_0)^T (\mathbf{x} - \boldsymbol{\mu}_0) \} \\ &= -\frac{1}{2\sigma^2} \{ (\mathbf{x}^T \mathbf{x} - 2\boldsymbol{\mu}_1^T \mathbf{x} + \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1) - (\mathbf{x}^T \mathbf{x} - 2\boldsymbol{\mu}_0^T \mathbf{x} + \boldsymbol{\mu}_0^T \boldsymbol{\mu}_0) \} + \ln \left\{ \frac{P(\mathbf{X}_1)}{P(\mathbf{X}_0)} \right\} \\ &= -\frac{1}{2\sigma^2} \{ -2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{x} + (\boldsymbol{\mu}_1^T \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T \boldsymbol{\mu}_0) \} + \ln \left\{ \frac{P(\mathbf{X}_1)}{P(\mathbf{X}_0)} \right\} \end{aligned}$$

Discriminant Function: $\mathbf{C}_1 = \mathbf{C}_0 = \sigma^2 \mathbf{I}$

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

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

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


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

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

Discriminant Function: $\mathbf{C}_1 = \mathbf{C}_0 = \mathbf{C}$

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

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

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

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

Discriminant Function: $\mathbf{C}_1 = \mathbf{C}_0 = \mathbf{C}$

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

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

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


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

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

Discriminant Function: $\mathbf{C}_1 \neq \mathbf{C}_0$

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

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

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

Discriminant Function: $\mathbf{C}_1 \neq \mathbf{C}_0$

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

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

Discriminant Function: $\mathbf{C}_1 \neq \mathbf{C}_0$

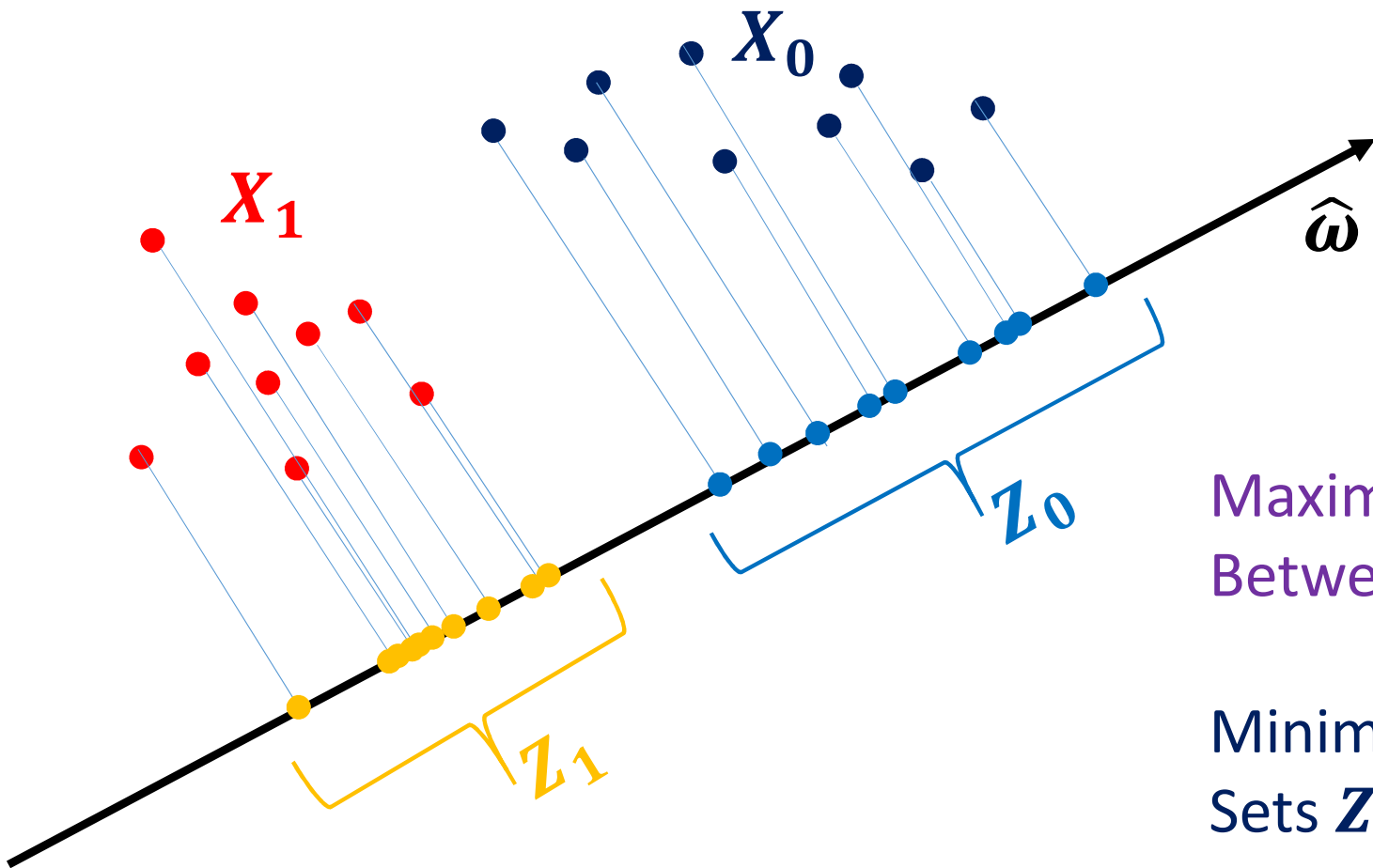
$$g(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$


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

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

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

Linear Discriminant Analysis (LDA)



Maximize Distance
Between Sets Z_0 and Z_1

Minimize Variances of
Sets Z_0 and Z_1

Recapitulation: Evaluating Mean and Variance

$$\mathbf{X} = \{\mathbf{x}_i : \mathbf{x}_i \in \mathbb{R}^D; i = 1, \dots, n\} \quad \mathbf{p}^T \mathbf{p} = 1$$

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

$$\gamma = \text{Mean}(\mathbf{Z}) = \frac{1}{n} \sum_i^n z_i \quad \boldsymbol{\mu} = \text{Mean}(\mathbf{X}) = \frac{1}{n} \sum_i^n \mathbf{x}_i$$

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

Recapitulation: Evaluating Mean and Variance

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

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

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

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

Definitions

$$\mathbf{X}_0 = \{\mathbf{x}_i^0 : \mathbf{x}_i^0 \in \mathbb{R}^D; i = 1, \dots, n_0\}$$

$$\mathbf{X}_1 = \{\mathbf{x}_j^1 : \mathbf{x}_j^1 \in \mathbb{R}^D; j = 1, \dots, n_1\}$$

$$\mathbf{Z}_0 = \{z_i^0 : z_i^0 = \boldsymbol{\omega}^T \mathbf{x}_i^0; \mathbf{x}_i^0 \in \mathbf{X}_0\}$$

$$\mathbf{Z}_1 = \{z_j^1 : z_j^1 = \boldsymbol{\omega}^T \mathbf{x}_j^1; \mathbf{x}_j^1 \in \mathbf{X}_1\}$$

$$y(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathbf{X}_1 \\ 0, & \mathbf{x} \in \mathbf{X}_0 \end{cases}$$

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

$$\boldsymbol{\mu}_0 = \text{Mean}(\mathbf{X}_0); \mathbf{C}_0 = \text{Cov}(\mathbf{X}_0)$$

$$m_0 = \text{Mean}(\mathbf{Z}_0); v_0 = \text{Var}(\mathbf{Z}_0)$$

$$\boldsymbol{\mu}_1 = \text{Mean}(\mathbf{X}_1); \mathbf{C}_1 = \text{Cov}(\mathbf{X}_1)$$

$$m_1 = \text{Mean}(\mathbf{Z}_1); v_1 = \text{Var}(\mathbf{Z}_1)$$

Problem Formulation: Step #1

Maximize Distance between Sets Z_0 and Z_1

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

$$\begin{aligned}(m_1 - m_0)^2 &= (\omega^T \mu_1 - \omega^T \mu_0)(\mu_1^T \omega - \mu_0^T \omega) \\ &= \omega^T (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T \omega\end{aligned}$$

Between-Class (X_0 and X_1) Scatter Matrix

$$S_B = (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T$$

Problem Formulation: Step #2

Minimize Net Variances of Sets Z_0 and Z_1

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

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

Net Within-Class (X_0 and X_1) Scatter Matrix

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

LDA: Objective

Maximize: $\omega^T S_B \omega$

Minimize: $\omega^T S_W \omega$

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



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

Recapitulation: KKT Conditions

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

Objective Function: $g(\mathbf{x}; \alpha_1, \dots, \alpha_n) = f(\mathbf{x}) + \sum_{i=1}^n \alpha_i h_i(\mathbf{x})$

KKT Conditions: $\nabla_{\mathbf{x}} g(\mathbf{x}; \alpha_1, \dots, \alpha_n) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^n \alpha_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0$

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

Objective Function: $g(\mathbf{x}; \alpha_1, \dots, \alpha_n) = -f(\mathbf{x}) + \sum_{i=1}^n \alpha_i h_i(\mathbf{x})$

KKT Conditions: $\nabla_{\mathbf{x}} g(\mathbf{x}; \alpha_1, \dots, \alpha_n) = -\nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^n \alpha_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0$

LDA: Solution

$$g(\boldsymbol{\omega}; \lambda) = -\boldsymbol{\omega}^T \mathbf{S}_B \boldsymbol{\omega} + \lambda(\boldsymbol{\omega}^T \mathbf{S}_W \boldsymbol{\omega} - 1)$$

Application of KKT Conditions

$$\nabla_{\boldsymbol{\omega}} g(\boldsymbol{\omega}; \lambda) = -2\mathbf{S}_B \boldsymbol{\omega} + \lambda(2\mathbf{S}_W \boldsymbol{\omega}) = 0$$



$$2\mathbf{S}_B \boldsymbol{\omega} = 2\lambda \mathbf{S}_W \boldsymbol{\omega}$$



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

LDA: Solution

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

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

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

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



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

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

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 4.1429 \\ 4.0000 \end{pmatrix}$$

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 1.7143 \\ 6.0000 \end{pmatrix}$$

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = \begin{pmatrix} 2.4286 \\ -2.0000 \end{pmatrix}$$

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

Example: Compute Covariances

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

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

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

Example: LDA

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



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

Comparison: LDA & PCA Directions

$$\begin{array}{c} \text{Best Principal} \\ \text{Component} \\ C \rightarrow \begin{pmatrix} \boxed{-0.7995} & 0.6007 \\ \boxed{-0.6007} & \boxed{-0.7995} \\ \boxed{9.0882} & \boxed{2.6924} \end{pmatrix} \\ \text{Eigen Values} \end{array}$$

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

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