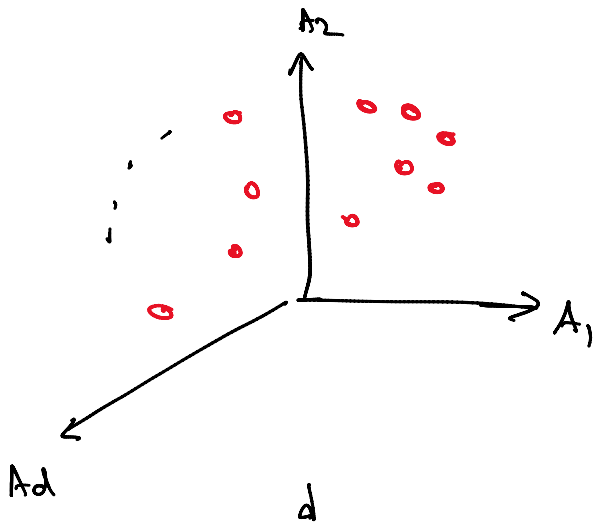
Lecture 5

Monday, September 11, 2023 10:02 AM



$\mathbb{R}^d$

$\hat{\mu} \equiv$ mean vector $\in \mathbb{R}^d$

$\hat{\sigma}_i^2 \equiv$ variance for A_i

$\hat{\sigma}_{ij}^1 \equiv$ Covariance (A_i, A_j)

$\hat{\Sigma} \equiv$ Cov matrix $\in \mathbb{R}^{d \times d}$

$d \times d$

$\hookrightarrow$ Positive semi-definite

real & non-negative eigenvalues

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$

$\vec{u}_1 \quad \vec{u}_2 \qquad \vec{u}_d$

$$\hat{\Sigma} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$ example

Q: Input D is in d-dim

$\{A_1, A_2, \ldots, A_d\}$

$\mathbb{R}^d$

If D is full rank it must have d linearly independent basis vectors $\quad d=3$

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$$\{\vec{e}_1, \underbrace{\vec{e}_2, \vec{e}_3, \ldots}, \vec{e}_d\}$$

orthonormal basis $\left\{ \begin{array}{l} \text{orthogonal} \\ \text{normal} \leftarrow \text{unit vectors} \end{array} \right.$

Q: Can we find a "better" basis in $\mathbb{R}^d$?

$$\hat{\Sigma} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \Longrightarrow \Sigma' = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\underbrace{\Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}}_{\text{original basis}} \Longrightarrow \underbrace{\Sigma' = \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}}_{\text{new basis}}$$

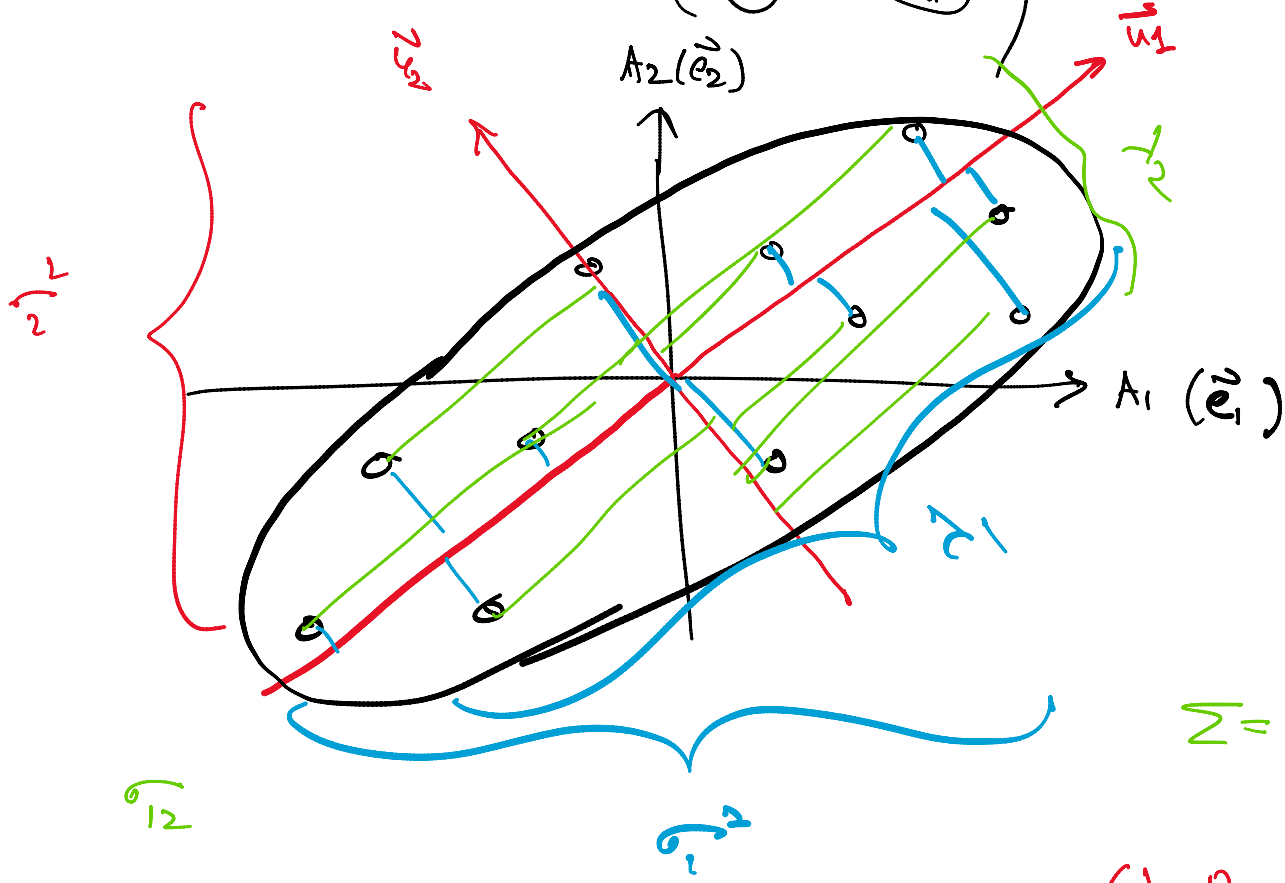
<u>Answer:</u> $D \leftarrow$ original basis $\equiv \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_d\}$

$\hookrightarrow \Sigma \equiv$ cov matrix $= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2d} \\ \vdots & & & \\ & & & \sigma_d^2 \end{pmatrix}$

after PCA

$\hookrightarrow$ new orthonormal basis $= \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_d\}$

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{pmatrix}$$



$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Cov matrix in new $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

Cov matrix in new basis! $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

PCA $\equiv$ Principal Component Analysis

$$\boxed{\Sigma u = \lambda u}$$

$\vec{u}_1 \quad - \lambda_1$

$\vec{u}_2 \quad - \lambda_2$

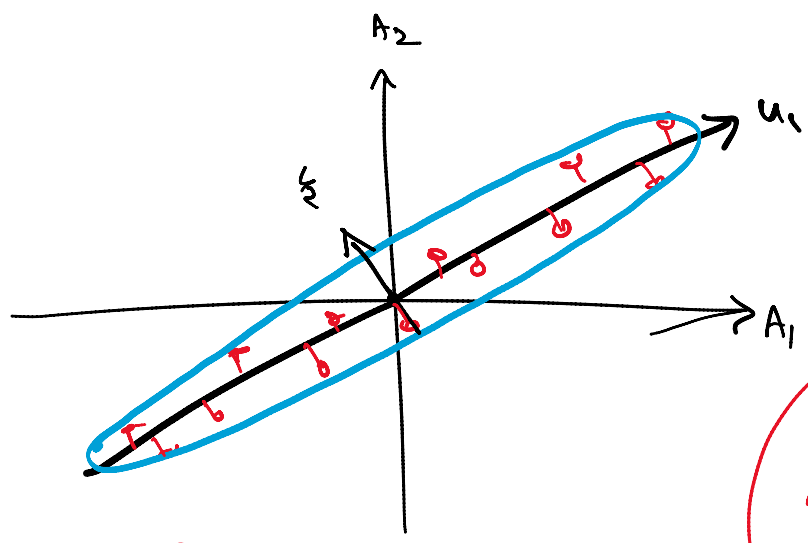
$\vdots$

$\vec{u}_d \quad - \lambda_d$

Choose top-k

basis vectors $\{$ Principal Components

Variances



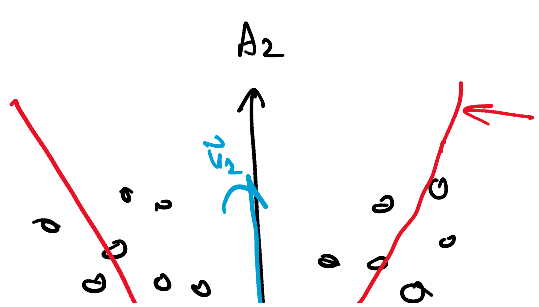
generative process

a) pick a point along $\vec{u}_1$

b) inject some $\varepsilon \geq 0$ noise

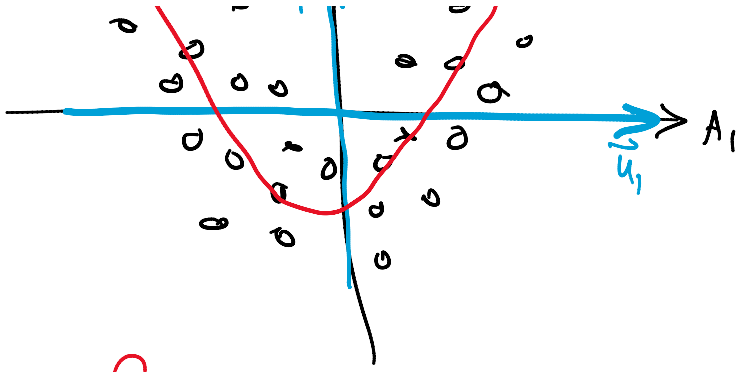
$\{A_1, A_2\} \in \mathbb{R}^2$

$\vec{u}_1 \in \mathbb{R}$ is a better representation of D



best curve to min MSE

mean Square



mean squared error

PCA $\equiv$ linear only!

$\vec{u}_1$ is a linear combination of $\{A_1\ A_2 \ldots A_d\}$

$$\vec{u}_1 = \underbrace{w_1 \vec{e}_1 + w_2 \vec{e}_2 + \cdots + w_d \vec{e}_d}$$

1st PC is <u>linear</u> combination of $\{\vec{e}_1\ \vec{e}_2 \ldots \vec{e}_d\}$

$A_1\ A_2 \ldots A_d$

$$D = \begin{array}{c} \begin{array}{cccc} A_1 & A_2 & \ldots & A_d \end{array} \\ \begin{bmatrix} 1 & 5 & \ldots & -1 \\ & & & \\ & & & \end{bmatrix} \end{array} \Longrightarrow D_q$$

quadratic

$$\begin{array}{c} A_1\ A_2 \ldots A_d\ \ A_1^2\ A_2^2 \ldots A_d^2\ \ A_1A_2\ A_1A_3 \ldots A_1A_d\ \ A_2A_3\ A_2A_4 \ldots \ldots A_{d-1}A_d \\ \begin{bmatrix} \quad & \quad & \quad & \quad & \quad & \quad & \quad & \quad & \quad \\ & & & - & - & - & & & \\ & & & & & & & & \end{bmatrix} \end{array}$$

$$A_1\ A_2\ A_3\ A_1^2\ A_2^2\ A_3^2\ A_1A_2\ A_1A_3\ A_1A_4$$

	A_1	A_2	A_3
x_1^T	1	2	-1

$\xRightarrow{\phi}$

A_1	A_2	A_3	A_1^2	A_2^2	A_3^2	A_1A_2	A_1A_3	A_2A_3
1	2	-1	1	4	1	2	-1	-2

D_q

$\phi(\vec{x_1}) = \vec{x_1}'$

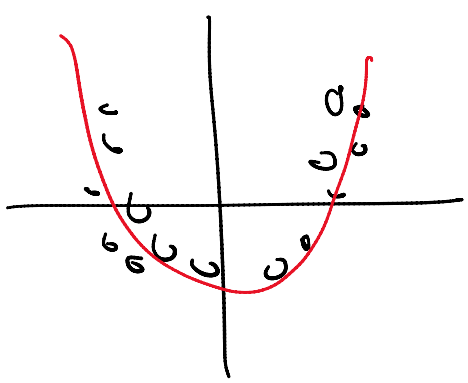
in new "quadratic" space

$\Downarrow$

perform PCA on this space

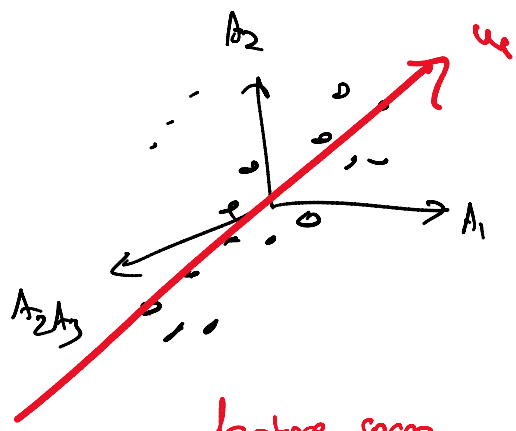
$\vec{u_1}'$ is a linear combination of

$\{A_1, A_2, A_3, A_1^2, A_2^2, A_3^2, A_1A_2, A_1A_3, A_2A_3\}$



Input space

$\Longrightarrow$



feature space

$$\boxed{\Sigma \vec{u} = \lambda \vec{u}}$$

d different equations

$$\begin{cases} \Sigma \vec{u_1} = \lambda_1 \vec{u_1} \\ \Sigma \vec{u_2} = \lambda_2 \vec{u_2} \end{cases}$$

$$\Sigma \vec{u}_2 = \lambda_2 \vec{u}_2$$

$$\vdots$$

$$\Sigma \vec{u}_d = \lambda_d \vec{u}_d$$

group all $\vec{u}$'s together

$$\Sigma \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_d \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix}$$

$d \times d$ $\qquad \underbrace{U}_{d \times d} \qquad U \qquad \Lambda$

$$\Sigma U = U \Lambda$$

← matrix equation to represent the d individual equations

all cols are eigenvectors

U is an orthogonal matrix ← $\boxed{U^{-1} = U^T}$

ortho-normal

$$\Sigma U = U \Lambda$$

$$\Sigma \cancel{U U^T} = U \Lambda U^T$$

Connection to Singular ← $\boxed{\Sigma = \underline{U} \Lambda \underline{U}^T}$ ← spectral decomposition

Singular Value Decomposition

$$\text{Cov} = \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & & \lambda_d \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_d^T \end{bmatrix}$$

$$\Sigma = U \quad \Lambda \quad U^T$$

decomposition

$\vec{u}_1 \in \mathbb{R}^d$ $(d \times 1)$

$\vec{u}_1^T : 1 \times d$

$$\boxed{\Sigma = \lambda_1 \vec{u}_1 \vec{u}_1^T + \lambda_2 \vec{u}_2 \vec{u}_2^T + \cdots + \lambda_d \vec{u}_d \vec{u}_d^T}$$

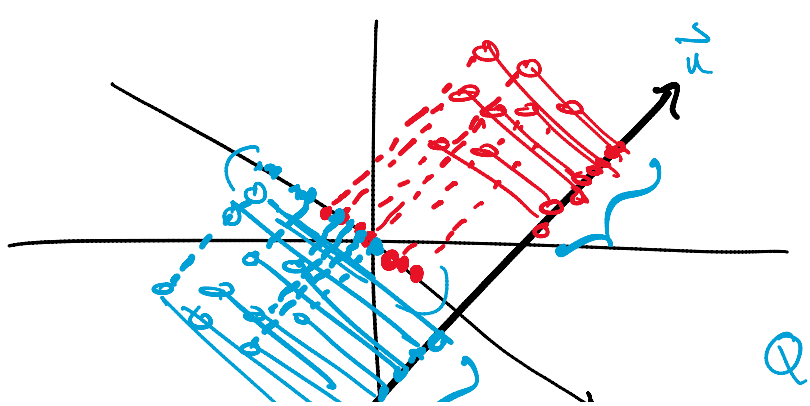
$$\underset{d}{\boxed{\Sigma}} = \lambda_1 \underset{d}{\boxed{u_1 u_1^T}}_{d} + \lambda_2 \underset{d}{\boxed{u_2 u_2^T}}_{d} + \cdots + \lambda_d \underset{d}{\boxed{u_d u_d^T}}_{d}$$

Linear Discriminant Analysis (LDA)

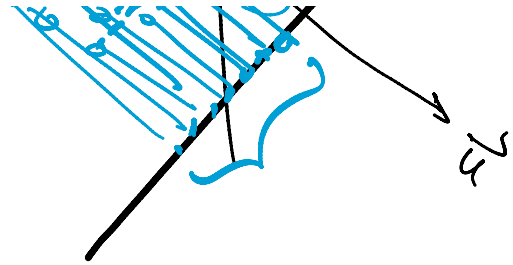
↳ supervision as <u>labels</u>

class = { o, o }

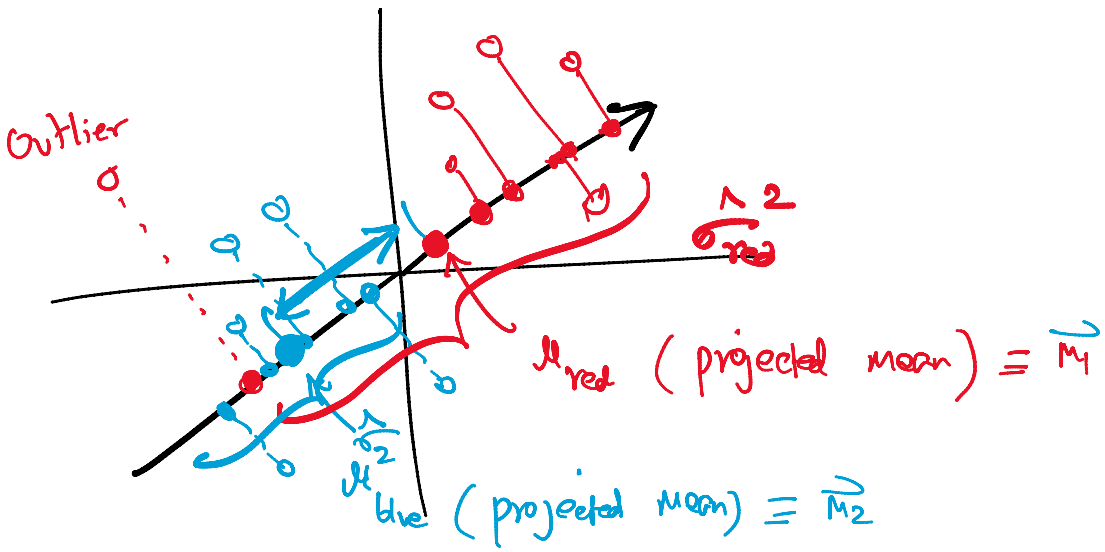
red, blue



Q: find a direction



Q: find a direction that $\boxed{\text{best separates}}$ the two groups



Objective 1: maximize $(\vec{m_1} - \vec{m_2})^2$

Projected mean vectors $\begin{cases} \vec{m_1} = \text{for class 1} \\ \vec{m_2} = \text{for class 2} \end{cases}$

Objective 2: minimize the sum of the variances

$$\sigma_1^2 + \sigma_2^2$$

$$\underbrace{s_1^2 + s_2^2}_{\text{Scatter values}}$$

$$s_1^2 = \sum_{j \in C_1} (a_j - m_1)^2$$

$$s_2^2 = \sum_{j \in C_2} (a_j - m_2)^2$$

scalar

where both a_j and m_i are projected values

$D \to D_1$ = all points from C_1

$D \to D_2$ = all points from C_2

C_1 & C_2 are the two classes

$|D_1| = n_1$ $\quad |D_2| = n_2$ $\quad |D| = n$

Q: find $\vec{u}$?

$\vec{x}_i \in D \longrightarrow$ Scalar projection onto $\vec{u}$

$$\frac{x_i^T \vec{u}}{\vec{u}^T \vec{u}}$$

$\vec{u}$ is unit vector

$$\equiv \underline{\underline{x_i^T \vec{u}}} = a_i$$

Scalar projection

$$m_1 = \frac{1}{n_1} \sum_{x_i \in D_1} a_i$$

$$m_2 = \frac{1}{n_2} \sum_{x_i \in D_2} a_i$$

$$\underline{I. \ (m_1 - m_2)^2}$$

maximize

$$\sigma_1^2 = \underbrace{\frac{1}{n_1} \sum_{x_i \in D_1} (a_i - m_1)^2}_{s_1^2}$$

$$\boxed{s^2 = \underbrace{\sum_{x_i \in D_1} (a_i - m_1)^2}}$$

$$\underbrace{\qquad\qquad}_{\text{Variance}} \qquad \boxed{\underbrace{\qquad\qquad}_{\text{Scatter}}}$$

$$s_2^2 = \sum_{x_i \in D_2} (a_i - m_2)^2$$

II: minimize $s_1^2 + s_2^2$

$$\max_{\vec{u}} \quad J = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \quad \begin{matrix}\uparrow \\ \downarrow\end{matrix}$$

Replace all m_i & s_i^2 with $\vec{x}_i$'s and $\vec{u}$

$$(m_1 - m_2)^2 = (\vec{u}^T \vec{\mu}_1 - \vec{u}^T \vec{\mu}_2)^2$$

$$m_1 = \frac{1}{n_1} \sum_{x_i \in D_1} \vec{u}^T \vec{x}_i = \vec{u}^T \left(\frac{1}{n_1} \sum_{x_i \in D_1} x_i \right) = \vec{u}^T \vec{\mu}_1$$

mean vector for D_1

$$(m_1 - m_2)^2 = \left\| u^T (\vec{\mu}_1 - \vec{\mu}_2) \right\|^2$$

$$= \vec{u}^T \boxed{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T} \vec{u}$$

$$= \vec{u}^T B \vec{u}$$

$(\vec{\mu}_1 - \vec{\mu}_2)$

Outer product of the diff vector with itself.

$$s_1^2 = \sum_{x_i \in D_1} (a_i - m_1)^2$$

$$= \ldots\ldots$$

$$= u^T \Big(\underbrace{\sum_{x_i \in D_1} (\vec{x}_i - \vec{\mu}_1)(x_i - \mu_1)^T}_{} \Big) u$$

$$= \vec{u}^T S_1 \vec{u}$$

$$S_1 = n_1 \cdot \Sigma_1$$

$$S_2 = n_2 \cdot \Sigma_2$$

$$J = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} = \frac{\vec{u}^T B \vec{u}}{\vec{u}^T (S_1 + S_2) \vec{u}}$$

$$\max_{\vec{u}} \quad J = \frac{\vec{u}^T B \vec{u}}{\vec{u}^T S \vec{u}}$$

$$S = S_1 + S_2$$

total scatter matrix

$$\boxed{\frac{\partial J}{\partial \vec{u}} = 0}$$

$$R\vec{u} = \lambda \vec{u}$$

$$\boxed{B\vec{u} = \lambda S \vec{u}}$$

$$\cancel{B\vec{u} = \lambda \vec{u}}$$

generalized eigenvalue problem

If S^{-1} exist,

$$S^{-1}(B\vec{u}) = \lambda S^{-1} S \vec{u}$$

$$\boxed{S^{-1}B}\,\vec{u} = \lambda \vec{u}$$

$\vec{u}$ solution is an eigenvector of $S^{-1}B$ matrix

$\vec{u}_1$ is actual solution

dominant eigenvector of $S^{-1}B$