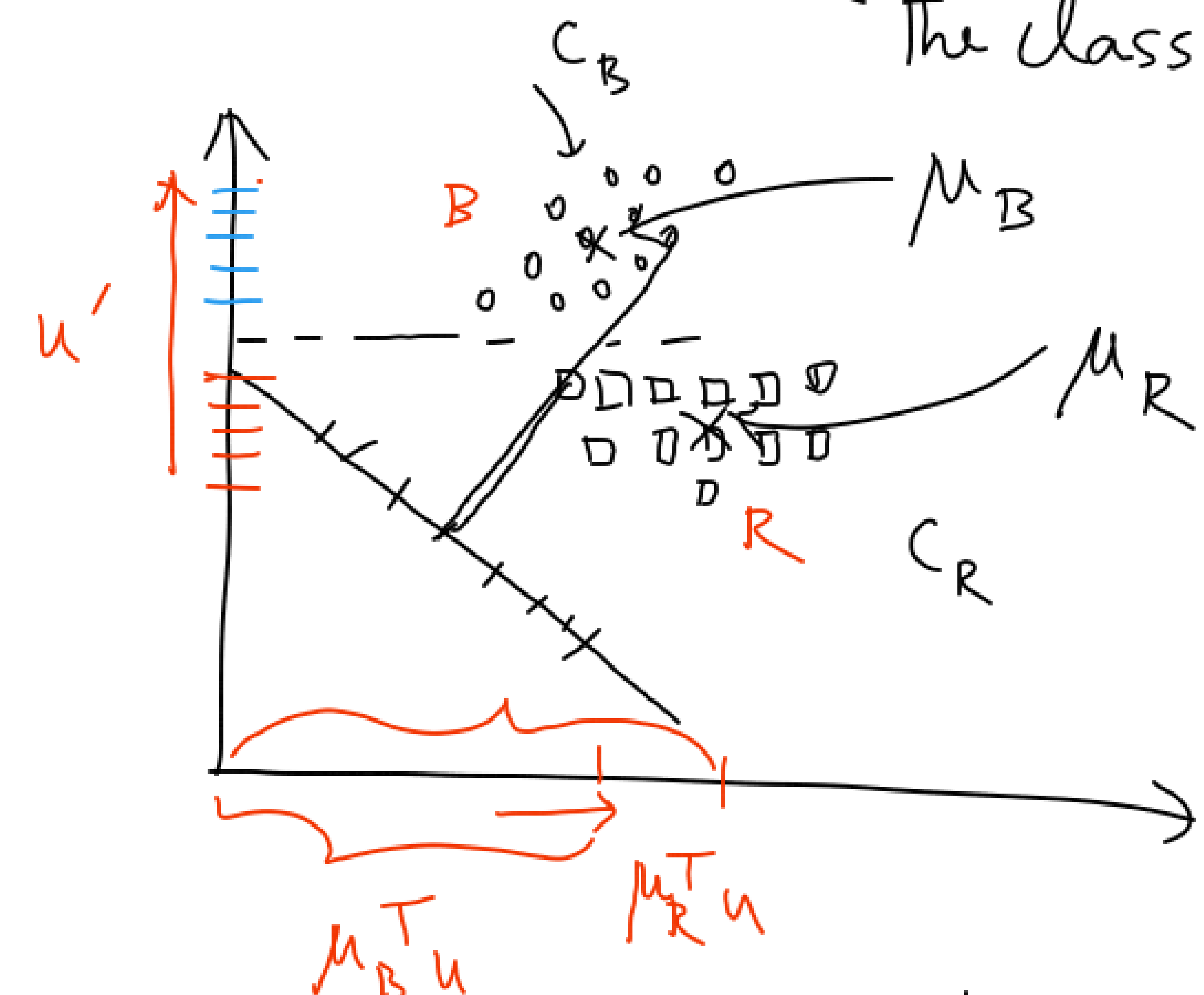
Lec 18: Dimensionality Reduction

Unsupervised: PCA $\rightarrow$ maximize the variance of the projected (does not use the class information)

Supervised:



Objective: (1) The means of the two classes are well separated.

$\rightarrow$ (2) The data in the same class are NOT well separated.

$$J(u) = |\mu_B^T u - \mu_R^T u|$$

Linear Discriminant Analysis

$(u^T x_i) u - (u^T \bar{x}) u$

$$= \frac{1}{n} \sum_i \left(u^T(x_i - \bar{x})\right)^2$$

$$= u^T S u$$

LDA: the variance within a class is called "scatter".

$x \in \mathbb{R}^d$

LDA: 2-class

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T$$

covariance matrix of class C_i

C_i: The data points belonging to class i

Variance of the projected data points of C_i on u

$$u^T S_i u \qquad S_W = S_1 + S_2$$

Sum of the within-class variances $= u^T S_1 u + u^T S_2 u$

$$= u^T S_W u$$

Variance between the means of two classes

$$(u^T \mu_1 - u^T \mu_2)^2 = u^T \underbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}_{} u$$

$$= u^T S_B u$$

S_B: between class covariance

LDA optimization problem:

$$J(u) = \frac{u^T S_B u}{u^T S_W u}$$

← this is scale invariant

$$\max J(u) \Rightarrow \left.\begin{aligned} &\max \; u^T S_B u \\ &\text{s.t. } u^T S_W u = 1 \end{aligned}\right\}$$

$$\mathcal{L}(\lambda, u) = -u^T S_B u + \lambda (u^T S_W u - 1)$$

(S_W is invertible)

$$\frac{\partial \mathcal{L}}{\partial u} = 0$$

$$\Rightarrow \; -2 S_B u + 2\lambda S_W u = 0$$

$$\Rightarrow \; \underbrace{S_W^{-1} S_B}\, u = \lambda u \;=\; S_B u = \lambda S_W u$$

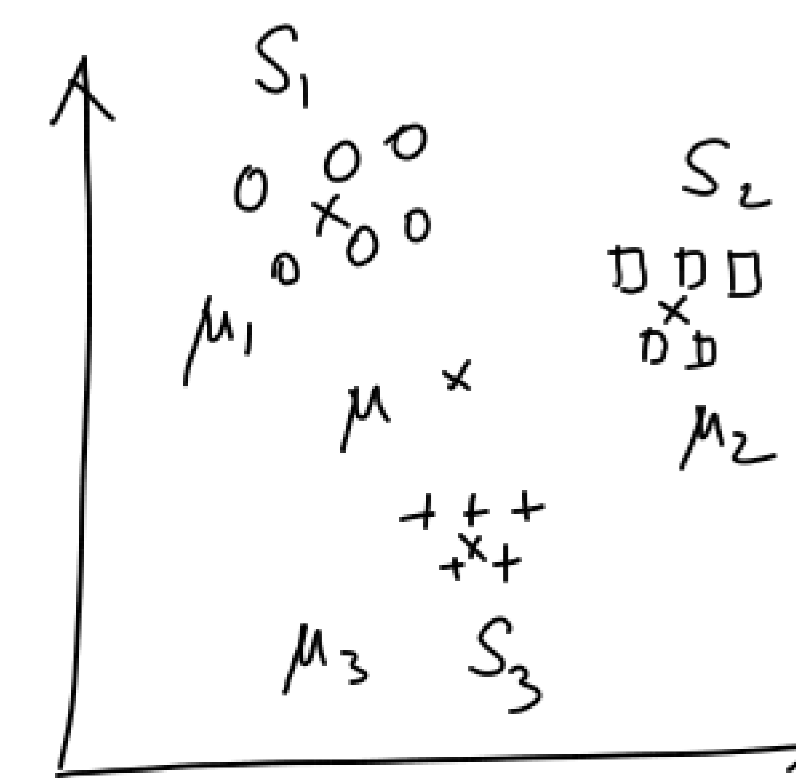
we should project the data points on a direction which is an eigenvector corresponding to the maximum eigenvalue

$$\boxed{\begin{aligned} &\max \; u^T(\lambda S_W u) = \lambda \\ &\text{s.t. } u^T S_W u = 1 \end{aligned}}$$

$$S_W^{-1} S_B = V \Sigma V^T$$

$\Rightarrow$ v_1 is the direction to project

LDA : $c > 2$ classes



$$S_W = S_1 + S_2 + S_3 + \cdots + S_c$$

$$S_B = \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^T$$

$|c_i|$ = number of data points in class i

$$\mu = \frac{1}{n} \sum_{i=1}^{c} n_i \mu_i$$

$$J(u) = \frac{u^T S_B u}{u^T S_W u}$$

$k \geq c$

Optimal solution $\Rightarrow$ $S_W^{-1} S_B u = \lambda u$

$$S_B = \begin{bmatrix} \underline{\sqrt{n_1}(\mu_1-\mu)} & \underline{\sqrt{n_2}(\mu_2-\mu)} & \cdots & \underline{\sqrt{n_c}(\mu_c-\mu)} \end{bmatrix} \begin{bmatrix} -\sqrt{n_1}(\mu_1-\mu)^T- \\ \vdots \\ -\sqrt{n_c}(\mu_c-\mu)^T- \end{bmatrix}$$

$c \ll d$ $\quad d \times c$

$\text{rank}(A) \leq c-1$

$$= A A^T$$

$$\text{rank}(AB) = \min(r(A), r(B))$$

$$\sum \gamma_i x_i = 0$$

if γ_i's are not all zeros

x_i's are linearly dependent

$$\sqrt{n_1} \cdot \sqrt{n_1}(\mu_1-\mu) + \cdots + \sqrt{n_c} \cdot \sqrt{n_c}(\mu_c-\mu) = 0$$

$\Rightarrow$ S_B is a low rank matrix

$\Rightarrow$ we can find at most $(c-1)$ discriminatory directions.

LDA Algorithm $(c>2)$

1. Compute the means of each class μ_i
2. Calculate S_W and S_B
3. Find top k eigenvectors corr to non-zero eigenvalues of $\underline{S_W^{-1} S_B}$

   $k \leq c-1$

   $u_1 \cdots u_k$

   Create $U = [\underline{u_1} \cdots \underline{u_k}]$
4. project x to $U^T x$.

Artificial Intelligence

reason

	Human side	Rational side
Thinking	NLP, Vision, automated reasoning ... ML	Logician's approach (complete information) can't handle uncertainties
Acting	Cognitive science – brain's functions	Agent based approach – single agent – multiple agent

Ref: Russell & Norvig

What is Rationality?

Making decisions with reason.

depends on:

1. performance measure
2. agent's prior knowledge about the environment and other agents
3. actions available to the agent
4. history (past states/actions)

Rationality

- ML → loss function (minimize)
- Robotics → Reinforcement learning
  → reward function
- Multi-agent systems → ≥ 2 agent
  → utility functions

Two player game :

A	B	C
-50 50	1 3	-10 20

- You choose one of the bins
- Opponent chooses a number from that bin
- You pm/utility is the number picked

Opponent is adversarial → B
random $\left(\frac{1}{2}, \frac{1}{2}\right)$ → C

Game Tree