CNU 데이터 분석 교육

10th lecture

"Linear Analysis-4"

2021 - 11 - 16

수학과 최명수

지난시간?

1. Minimization of loss function

   - Gradient method, Gird search, Calculus...

2. SST = SSE + SSR

   - On linear regression with RMSE loss function

오늘은 무엇을?

1. Linear discriminant analysis(LDA)

   -선형함수를 이용한 데이터 분류하기

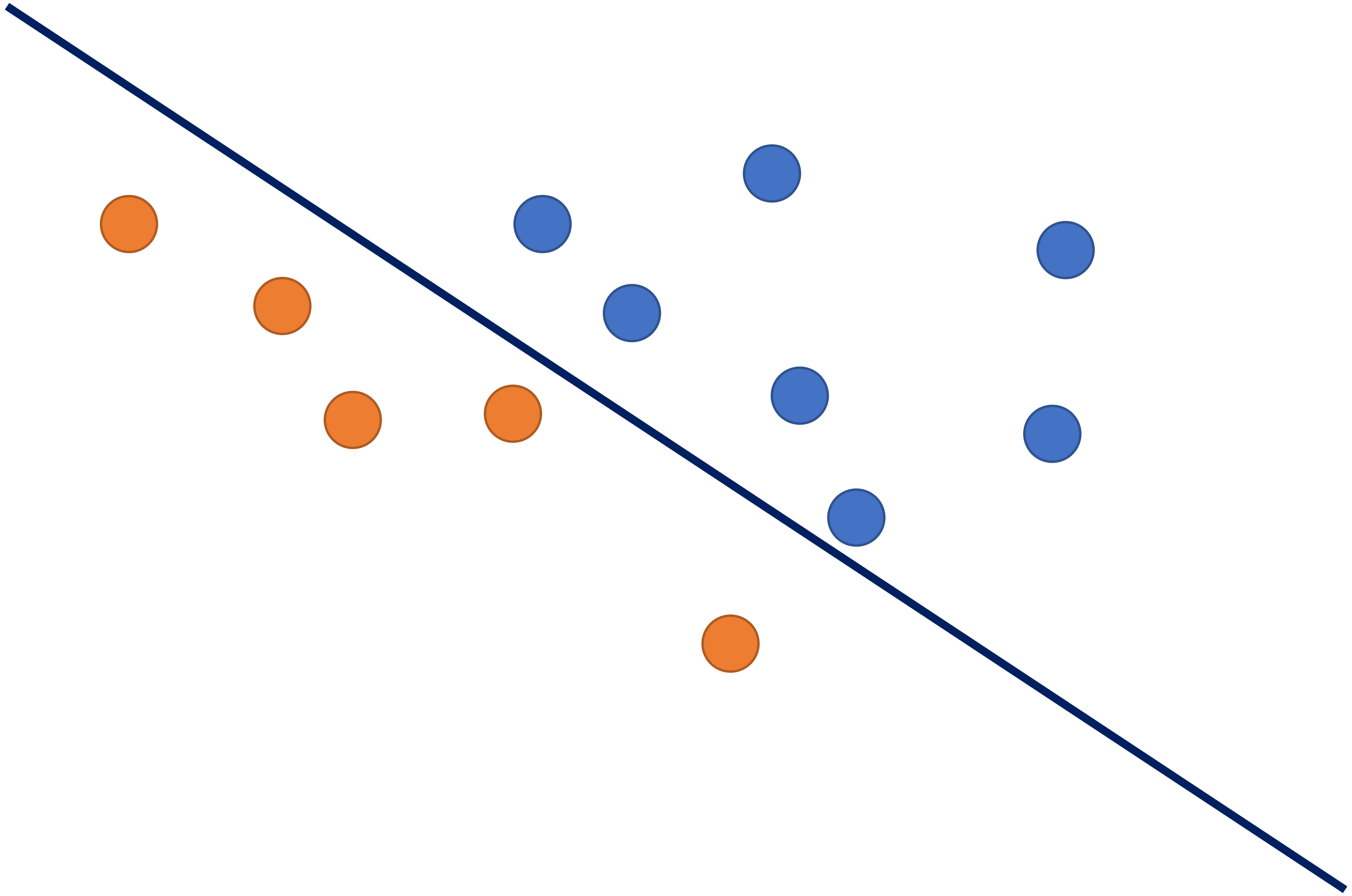
LDA

https://lazypenguins.com/where-bodies-of-water-with-different-colours-meet-around-the-world/

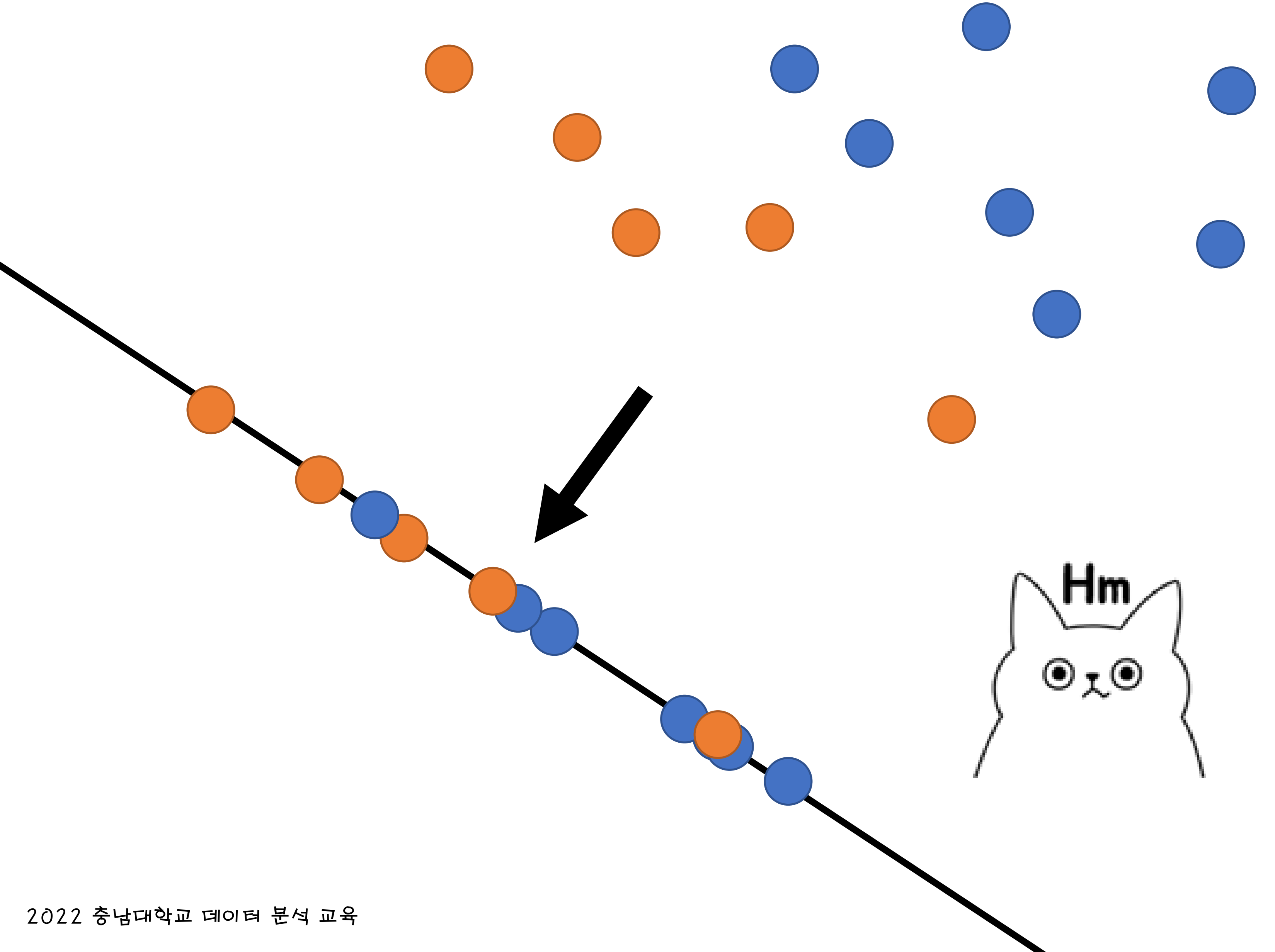
LDA

Sir Ronald Aylmer Fisher
(17 February 1890 – 29 July 1962)

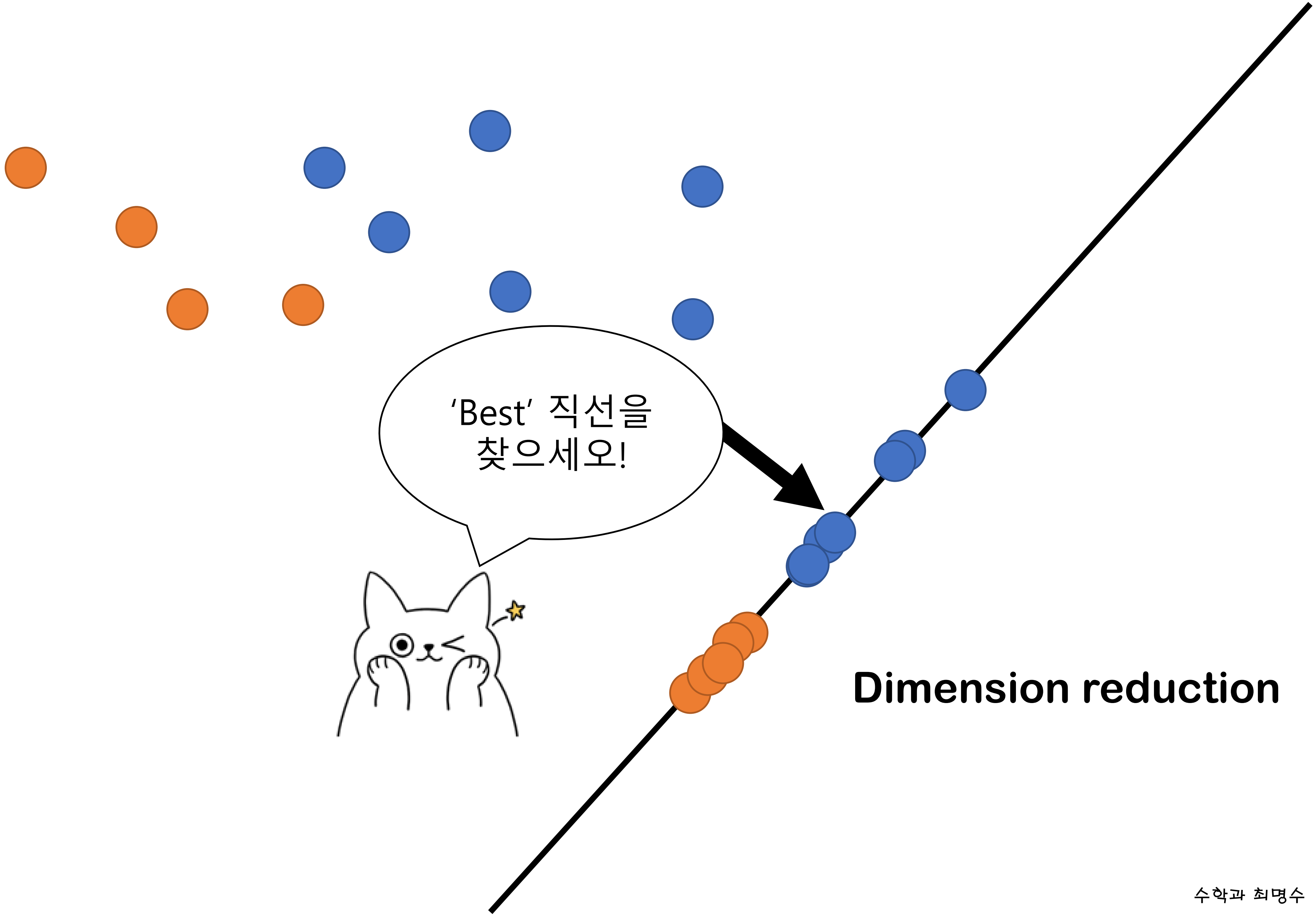
LDA



LDA



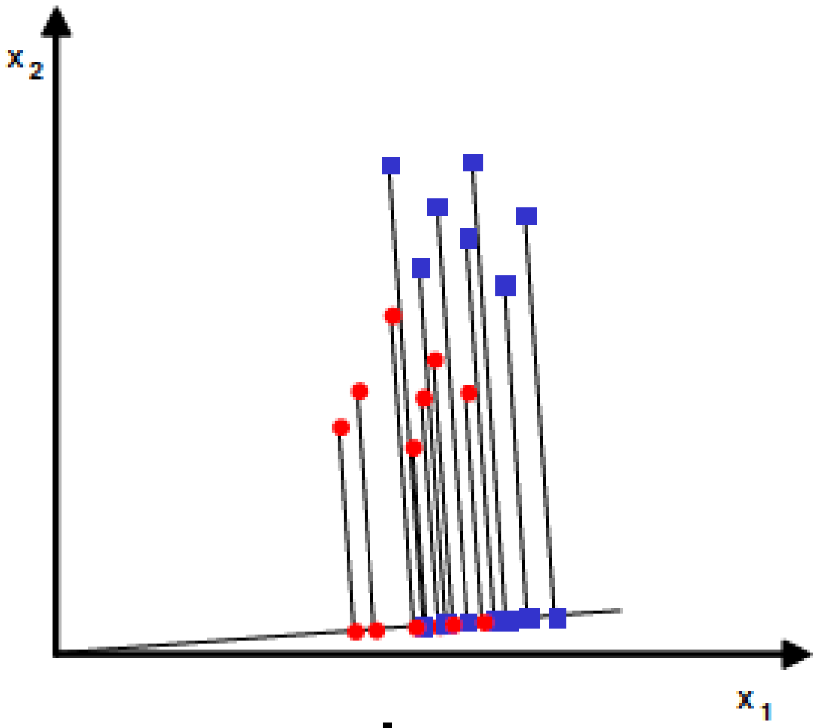
LDA



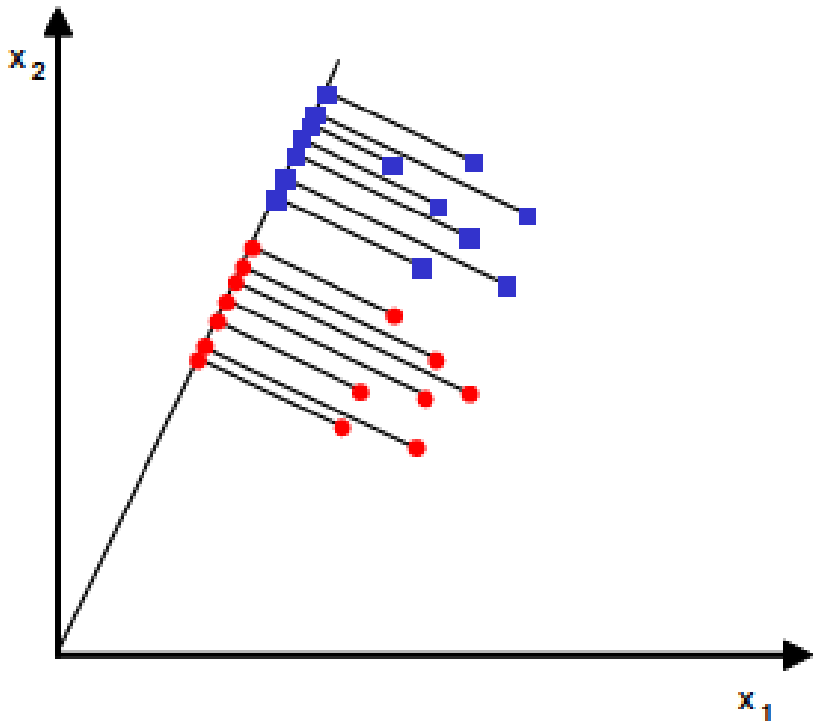
Dimension reduction

LDA

뭐가 최고인데???



$$|\mu_1 - \mu_2|$$

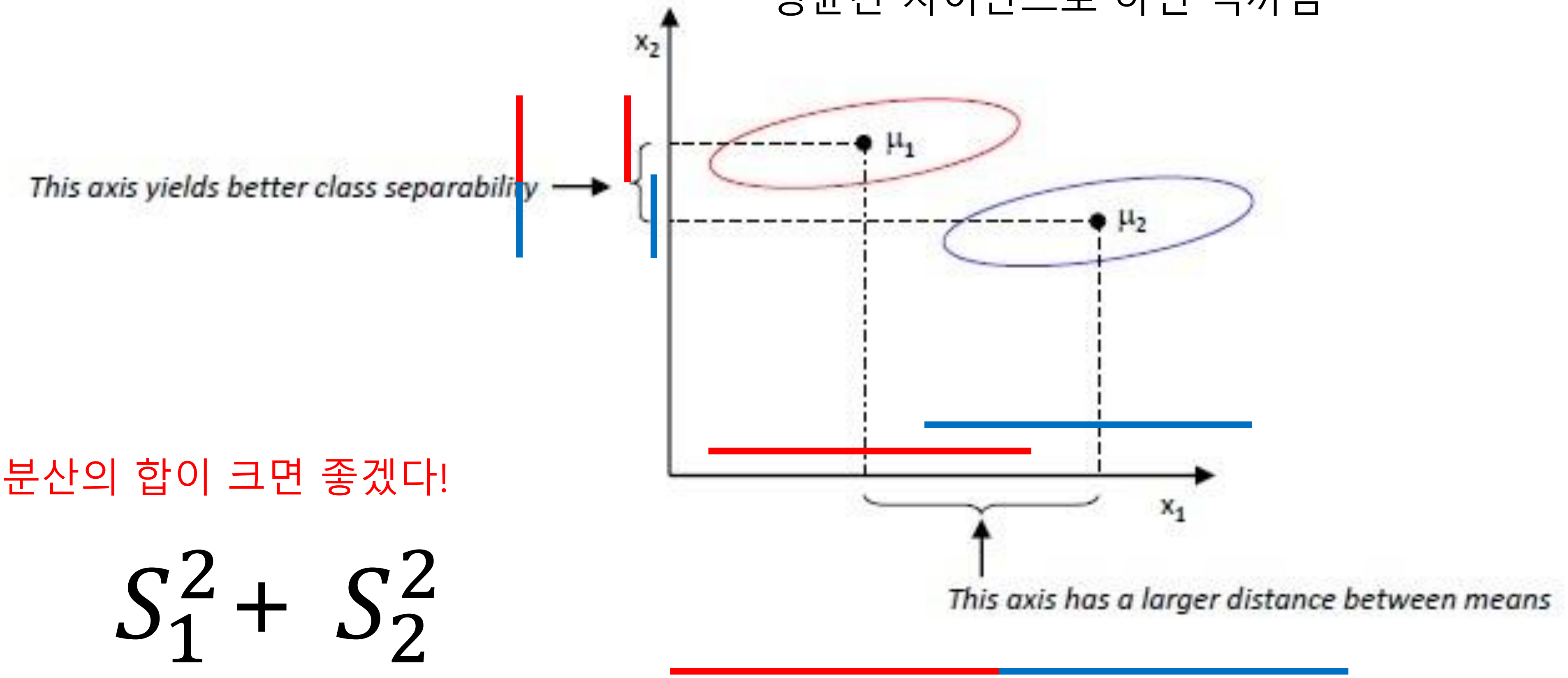


평균간의 차이가 크면 좋겠다!

https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=wjddudwo209&logNo=80208395369

LDA

평균간 차이만으로 하면 억까임



분산의 합이 크면 좋겠다!

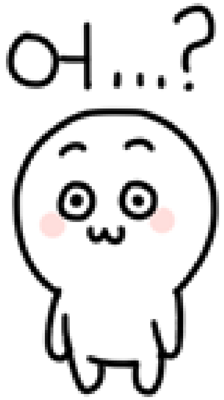
$$S_1^2 + S_2^2$$

https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=wjddudwo209&logNo=80208395369

LDA

$$Min_{\vec{w}} \ |\mu_1 \ - \mu_2|$$

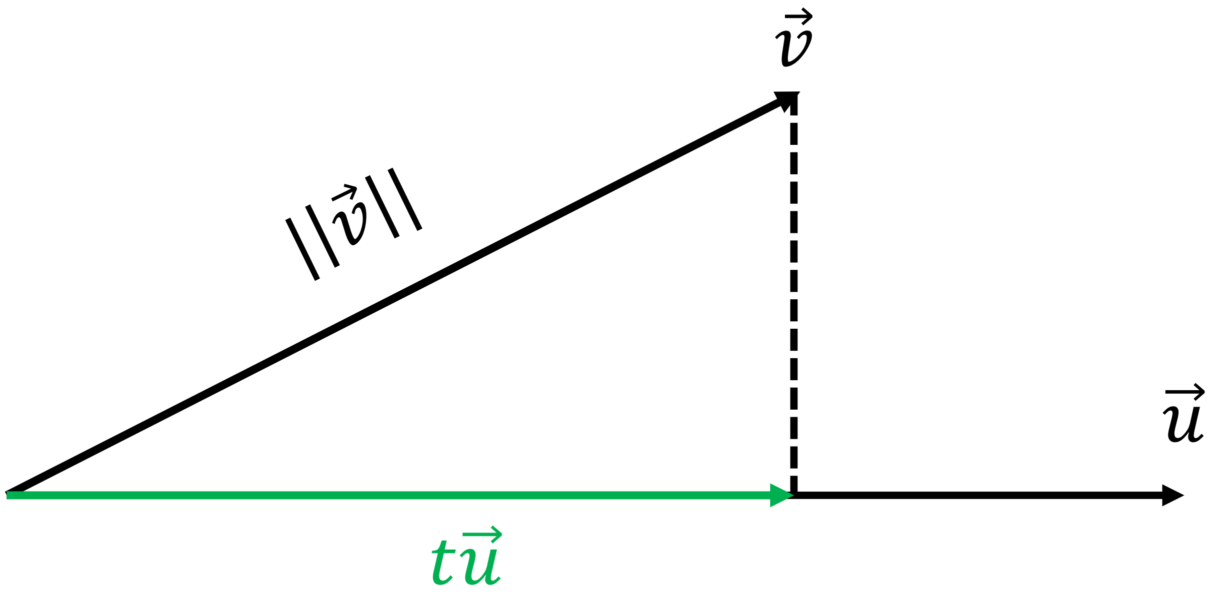
$$Max_{\vec{w}} \ \left|S_1^2 + S_2^2\right|$$



근데 우리의 데이터는 (x, y) 형태.. 즉 벡터형인데...?

$$J(\overrightarrow{w}) = \frac{|\mu_1 - \mu_2|}{S_1^2 + S_2^2}$$

LDA



Projection은 벡터의 내적!

$$proj_{\vec{u}}\vec{v} = \frac{\vec{u} \cdot \vec{v}}{\vec{u} \cdot \vec{u}}\vec{u} = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}|^2}\vec{u}$$

LDA

$$\vec{X}_i \ : \ \text{data}, \quad |\mathrm{A}| = n, \ |\mathrm{B}| = m, \ w^T : \text{vector}$$

$$\vec{M}_A = \frac{1}{n} \sum_{x_i \in A} \vec{X}_i \, , \quad \vec{M}_B = \frac{1}{m} \sum_{x_i \in B} \vec{X}_i$$

$$m_A = w^T M_A, \ \ m_B = w^T M_B \ : \ \text{Projected Mean}$$

LDA

$$x_i = w^T \vec{X}_i \ : \ \text{Projected data}$$

$$S_A = \sum_{x_i \in A} (x_i - m_A)^2, \quad S_B = \sum_{x_i \in B} (x_i - m_B)^2$$

$$Let \quad J(\vec{w}) = \frac{(m_A - m_B)^2}{S_A + S_B}$$

LDA

$$(m_A - m_B)^2 = (\vec{w}^T (M_A - M_B))^2$$

$$= \vec{w}^T (M_A - M_B)(M_A - M_B)^T \vec{w}$$

Let $S_M = (M_A - M_B)(M_A - M_B)^T$

Then $(m_A - m_B)^2 = \vec{w}^T S_M \vec{w}$

LDA

$$S_A = \sum_{x_i \in A} (x_i - m_A)^2 = \sum_{x_i \in A} (\vec{w}^T(\vec{X}_i - \vec{M}_A))^2$$

$$= \sum_{x_i \in A} \vec{w}^T(\vec{X}_i - \vec{M}_A)(\vec{X}_i - \vec{M}_A)^T \vec{w}$$

$$S_W = \sum_{x_i \in A} (\vec{X}_i - \vec{M}_A)(\vec{X}_i - \vec{M}_A)^T + \sum_{x_i \in B} (\vec{X}_i - \vec{M}_B)(\vec{X}_i - \vec{M}_B)^T$$

$$S_A + S_B = \vec{w}^T S_w \vec{w} \qquad \text{Then } \; J(\vec{w}) = \frac{\vec{w}^T S_M \vec{w}}{\vec{w}^T S_w \vec{w}}$$

LDA

$$\frac{d}{dw}J(\vec{w}) = \frac{d}{dw}\left(\frac{\vec{w}^T S_M \vec{w}}{\vec{w}^T S_w \vec{w}}\right)$$

$$= \frac{2S_M\vec{w}(\vec{w}^T S_w \vec{w}) - 2(\vec{w}^T S_M \vec{w})S_w\vec{w}}{(\vec{w}^T S_w \vec{w})^2}$$

$$2C_w S_M \vec{w} - 2C_M S_w \vec{w} = 0$$

$$S_w\vec{w} = \lambda S_M \vec{w} \quad \text{so, } \vec{w} = \lambda S_w^{-1} S_M \vec{w}$$

$$\vec{w} = \lambda S_w^{-1}(M_A - M_B)(M_A - M_B)^T \vec{w}$$

$$= k\lambda S_w^{-1}(M_A - M_B) \qquad \boxed{\therefore \vec{w} = S_w^{-1}(M_A - M_B)}$$

LDA

Input : $\vec{A}, \vec{B} \in R^2$

Calculate : $\overrightarrow{\mu_A},\ \overrightarrow{\mu_B},\ S_W$

Return : $\vec{w} = S_W^{-1}(\overrightarrow{\mu_A} - \overrightarrow{\mu_B})$

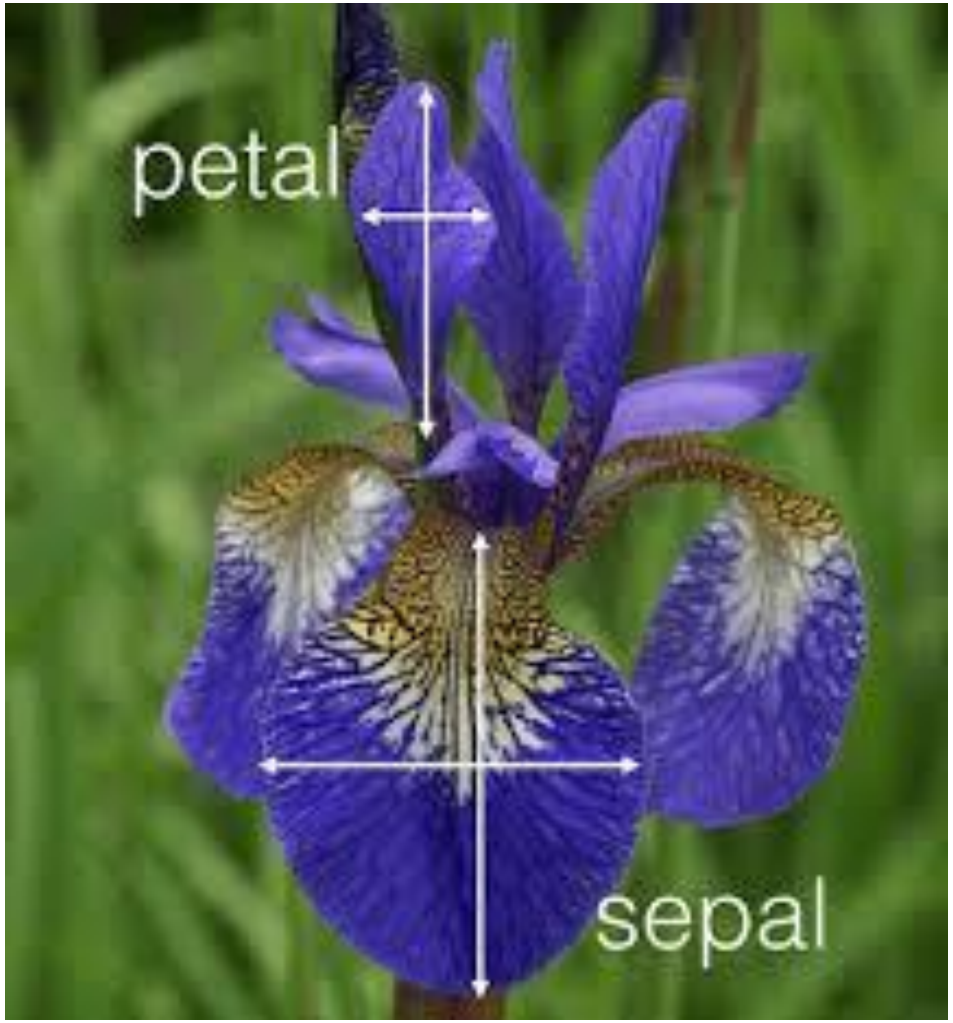
LDA

문제1

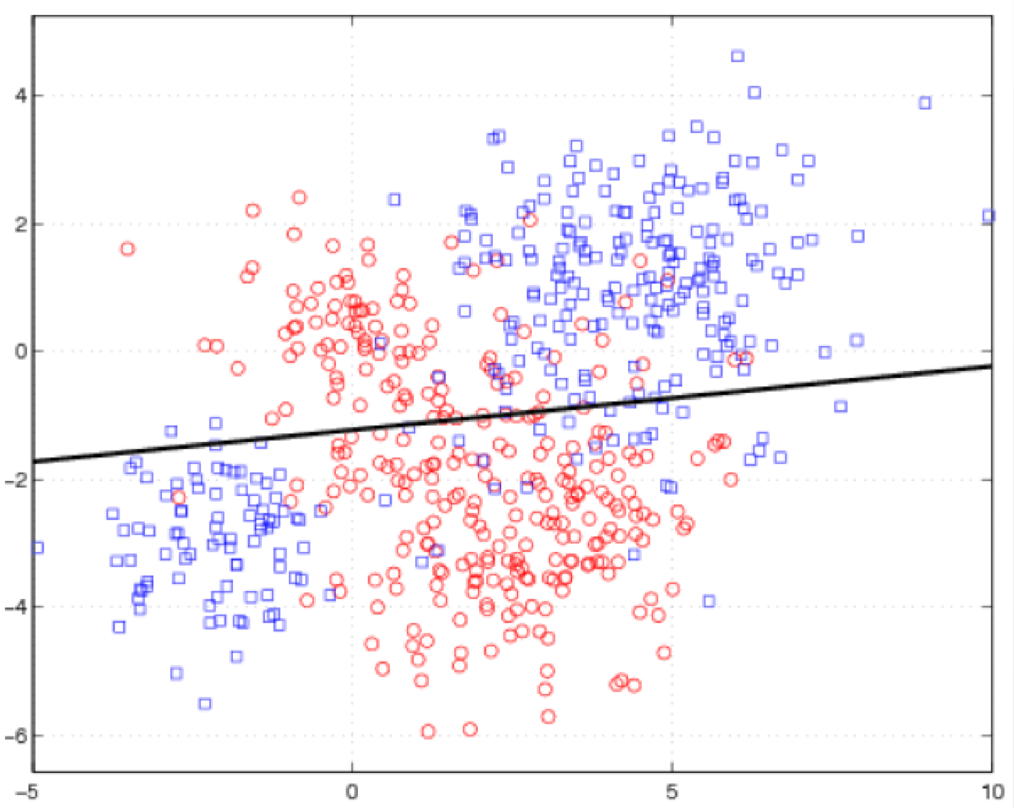
2차원 데이터를 이용하여 LDA를 하는 함수를 제작하라. 그리고 시각화 하여라.

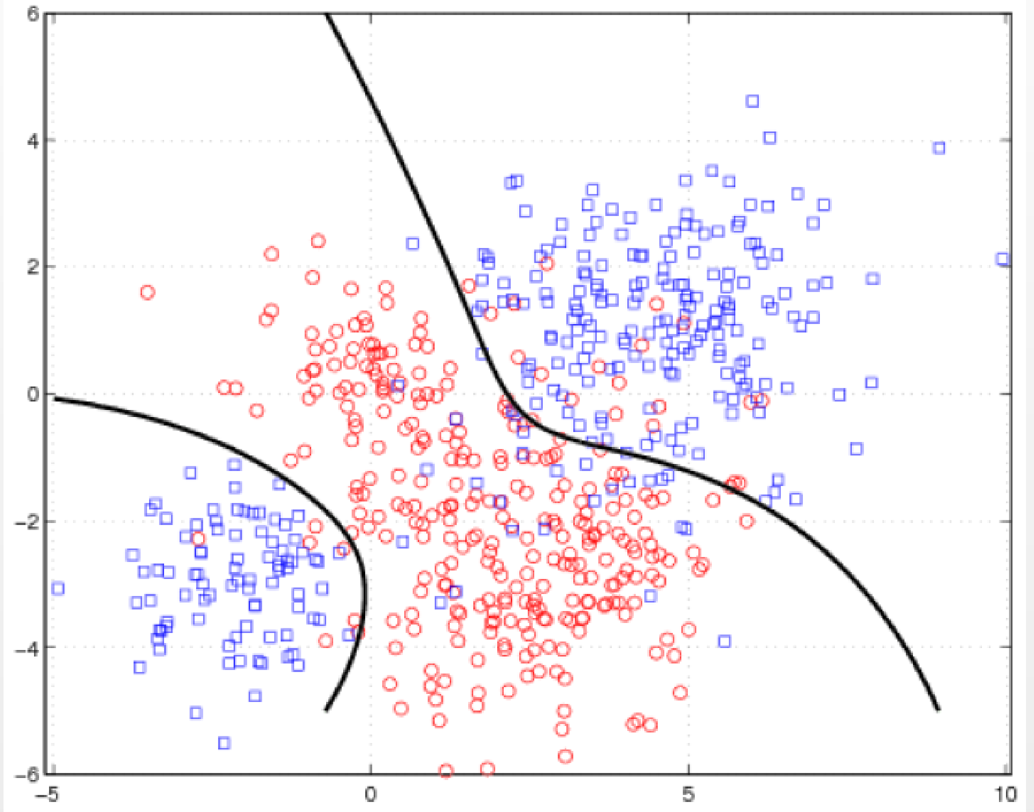
- 함수 이름 : LDA(X1, Y1, X2, Y2):
- Return : w, mu_A, mu_B

LDA



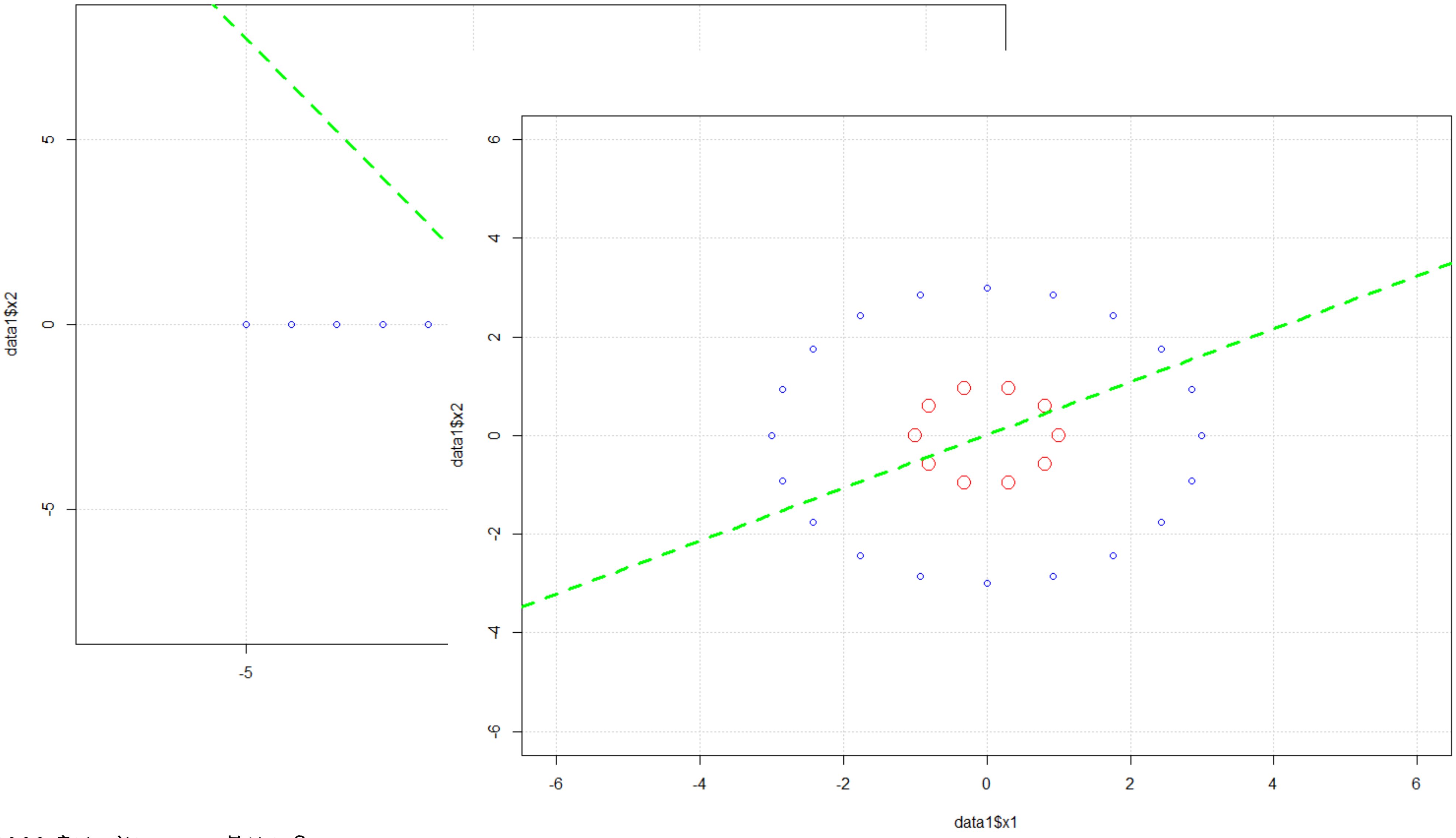
LDA 단점





Left: Decision boundaries by LDA. Right: Decision boundaries obtained by modeling each class by a mixture of two Gaussians.

LDA 단점



정리 요약

1. Linear discriminant analysis

   - 차원축소를 이용한 분석방법
   - 확실하게 구분되는 것은 정말 잘 구분한다.
   - 근데 세상의 대부분 문제들은 확실히 구별되지 않는 것들이다!

공지

1. 다음시간에는 인공신경망(Artificial neural network)기초에 대해 학습

끝

담에봅시당