R 교육 세미나

ToBig's 9기 최영제

LDA 차원축소

Unit 01 | LDA

LDA

Is LDA a dimensionality reduction technique or a classifier algorithm?

LDA 차원축소

주성분 분석법(PCA)은 데이터의 최적 표현의 견지에서 데이터를 축소하는 방법인데 반하여

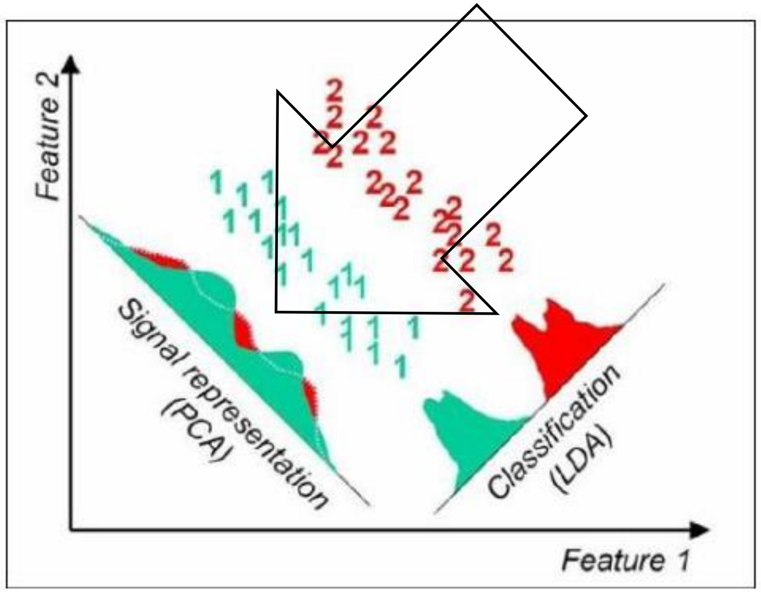
선형판별 분석법(LDA)은 데이터의 최적 분류의 견지에서 데이터를 축소하는 방법이라고 할 수 있다.

➔ (목적) 가능한 클래스간의 분별 정보를 최대한 유지시키면서 차원을 축소시키는 것

무슨말이지?

Unit 01 | LDA

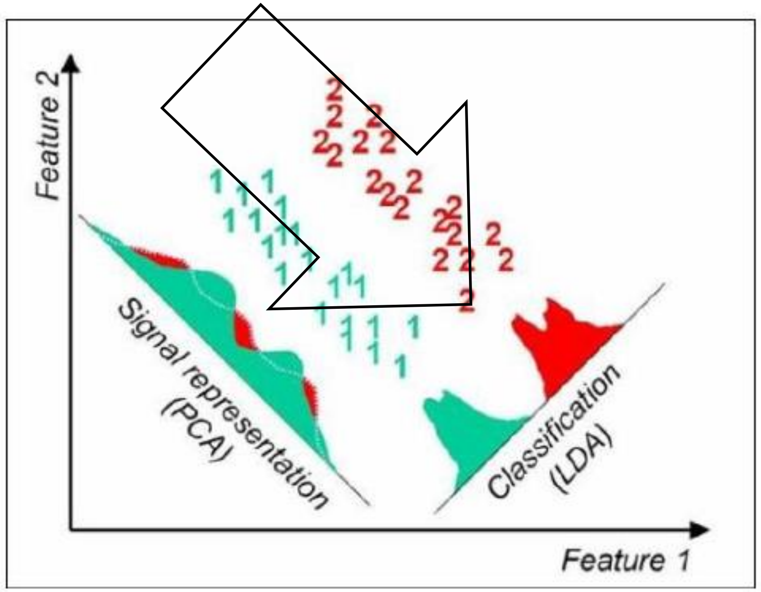
LDA



PCA, 분산을 가장 크게 바라보는 축이
차원 축소 시 정보의 손실이 적다는 개념(=최적표현의 견지)
그 과정에서 클래스를 고려하지 않는다

Unit 01 | LDA

LDA



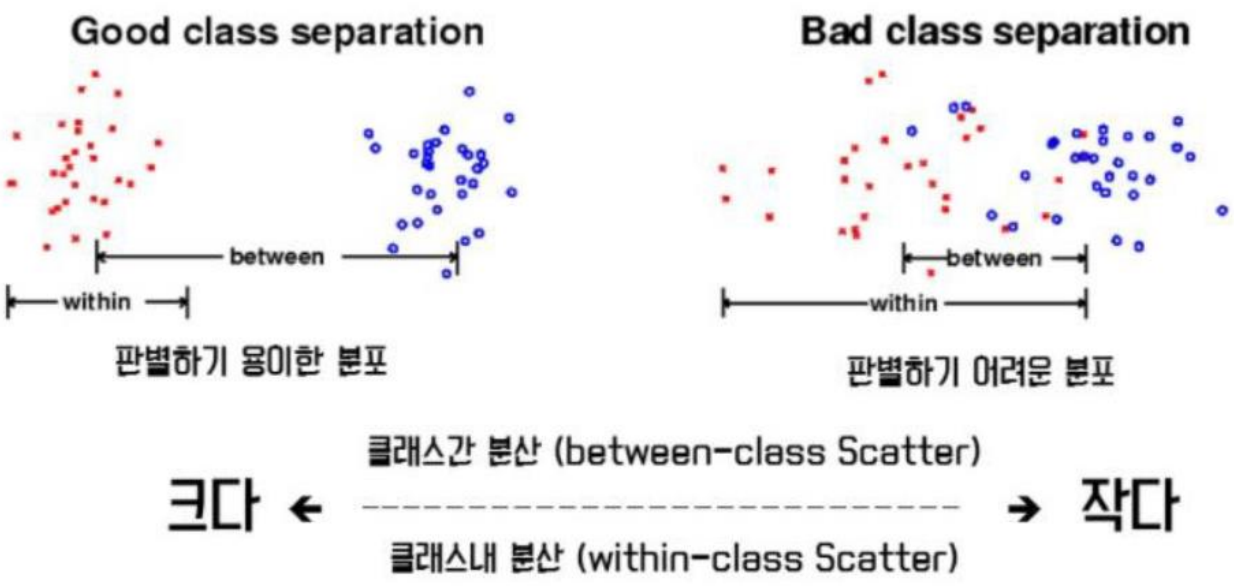
LDA, 특징 공간상에서 클래스 분리를 최대화하는
축으로 사상시켜 선형 부공간으로 차원 축소(=최적분리의 견지)

클래스간 분산(between-calss scatter)과
클래스내 분산(within-class scatter)의
비율을 최대화하는 방식으로 차원을 축소

->붉은 글씨 두개 기억해두세요!

Unit 01 | LDA

LDA



LDA 핵심은

클래스 간의 분산은 크게

= 클래스들은 서로 멀리(분자)

클래스 내의 분산은 작게

= 클래스 안에선 가깝게(분모)

$$J = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

LDA의 목적함수!

수식적으로 들어가봅시다!

이해 안가시는 분들은 SKIP하셔도 됩니다!

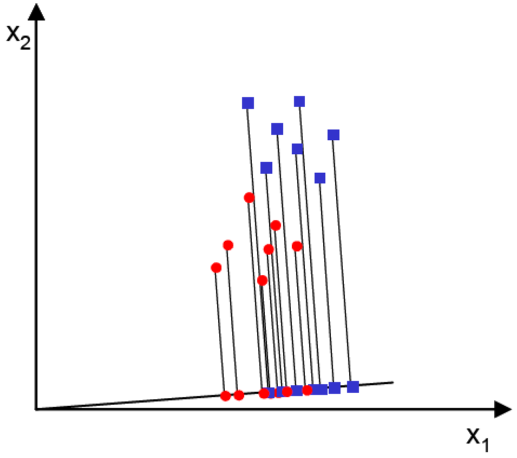
Unit 01 | LDA

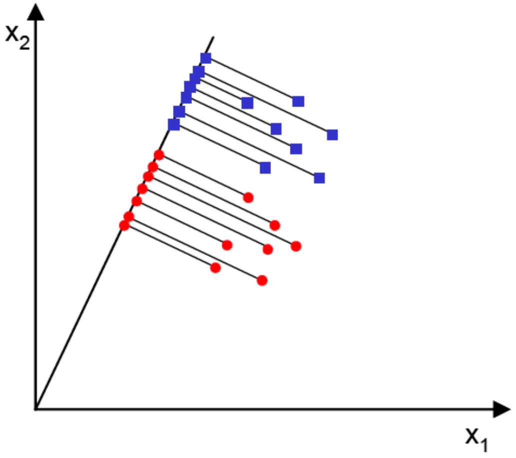
LDA

- D-차원 표본 데이터 집합 $X=\{x^{(1}, x^{(2}, \ldots, x^{(N}\}$ 가 주어진 경우, ω_1 클래스에 속하는 것이 N_1개 이고, ω_2 클래스에 속하는 것이 N_2개 일 때, x를 임의의 선을 따라서 사영하여 스칼라 y를 얻고자 한다.

$$\boxed{y = W^T x}$$

- 가능한 모든 선들 중에서 이러한 스칼라 값들의 분리를 최대화하는 것을 선택한다. (2차원의 경우를 예를 들면 다음과 같다.)

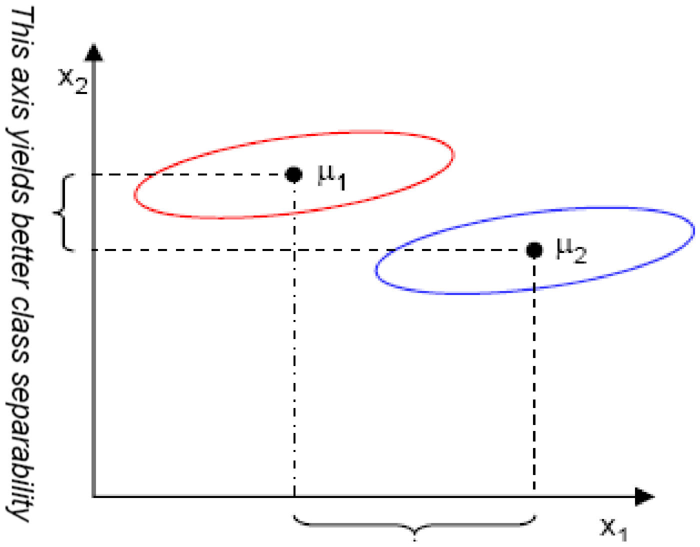




어느 사영을 취하는 것이 좋을 것인가?

Unit 01 | LDA

LDA



This axis has a larger distance between means

-> X축이 y축에 비하여 집단간 중심점의 거리가 먼 축이지만
두 집단을 구분하기엔 적합하지 않음을 알 수 있다

우선 두 집단간의 거리를 멀리 띄워보는 것으로 출발해보자

평균을 기준 척도로 하면, 각 클래스들의 x 와 y 에서의
평균백터는 다음과 같다.

$$\mu_i = \frac{1}{N_i}\sum_{x\in\varpi_i} x$$

$$\tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\varpi_i} y = \frac{1}{N_i}\sum_{x\in\varpi_i} w^T x = w^T \mu_i$$

사영된 데이터들의 **중심(평균)간의 거리**를 목적 함수로 선택하면,

$$J(w) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T(\mu_1 - \mu_2)\right|$$

평균만을 고려하면, 클래스 안에서의
표준편차가 고려되지 않으므로 좋은 척도가 아니다!!

Unit 01 | LDA

LDA

Fisher 에 의해서 제안된 방법은 클래스간(between-class)의 차이를 최대화 시키는것에 더하여 **클래스내(within-class)의 스캐터로 정규화한 평균들 간의 차이**로 표현된 **함수를 최대화**시키는 것이다.

각 클래스들에 대하여 스캐터(공분산과 같은 개념)는 다음과 같이 주어지며

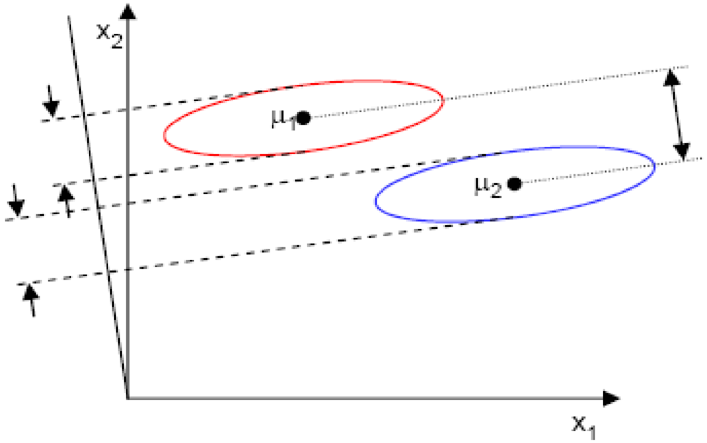
$$\widetilde{S}_i^2 = \sum_{y\in\omega_i}(y-\widetilde{\mu}_i)^2 = \sum_{y\in\omega_i}(y-\widetilde{\mu}_i)(y-\widetilde{\mu}_i)^T$$

각 클래스들의 x 와 y 에서의 평균백터

$$\mu_i = \frac{1}{N_i}\sum_{x\in\varpi_i} x$$

$$\widetilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\varpi_i} y = \frac{1}{N_i}\sum_{x\in\varpi_i} w^T x = w^T \mu_i$$

사영 표본들의 **클래스내 분산(within-class scatter)** ➔ $(\widetilde{S}_1^2 + \widetilde{S}_2^2)$



Fisher의 선형판별은 다음의 **목적함수**를 최대화 하는 선형함수 $w^T x$ 에 해당한다.

$$J(\mathbf{w}) = \frac{|\widetilde{\boldsymbol{\mu}}_1 - \widetilde{\boldsymbol{\mu}}_2|^2}{\widetilde{\mathbf{S}}_1^2 + \widetilde{\mathbf{S}}_2^2}$$

→ 최대

→ 최소

따라서, Fisher의 선형판별식은 동일한 클래스의 표본들은 인접하게 사영이 취해지고, 동시에 클래스 간의 사영은 중심이 가능한 멀리 떨어지게 하는 변환행렬(w)를 찾아내는 것이다.

Unit 01 | LDA

LDA

최적의 사영 w 를 구하기 위해서는 J(w) 를 w 에 대한 함수로 표현해야 한다.
다차원 특징 공간에서 스캐터(scatter) 행렬은 사영 상에서 분산과 동일한 형태

$$\mathbf{S}_i^2 = \sum_{\mathbf{x}\in\omega_i}(\mathbf{x}-\boldsymbol{\mu}_i)(\mathbf{x}-\boldsymbol{\mu}_i)^T$$

클래스내 스캐터(within-class scatter) 행렬 S_w ➔ $\mathbf{S}_1^2 + \mathbf{S}_2^2 = \mathbf{S}_w$

사영된 y의 스캐터를 특징벡터 x의 스캐터 행렬의 함수로 다음과 같이 표현된다.

$$\begin{aligned}
\tilde{\mathbf{S}}_i^2 &= \sum_{\mathbf{y}\in\omega_i}(\mathbf{y}-\tilde{\boldsymbol{\mu}}_i)(\mathbf{y}-\tilde{\boldsymbol{\mu}}_i)^T \\
&= \sum_{\mathbf{x}\in\omega_i}(\mathbf{w}^T\mathbf{x}-\mathbf{w}^T\boldsymbol{\mu}_i)(\mathbf{w}^T\mathbf{x}-\mathbf{w}^T\boldsymbol{\mu}_i)^T \\
&= \sum_{\mathbf{x}\in\omega_i}\mathbf{w}^T(\mathbf{x}-\boldsymbol{\mu}_i)(\mathbf{x}-\boldsymbol{\mu}_i)^T\mathbf{w} \\
&= \mathbf{w}^T\mathbf{S}_i^2\mathbf{w}
\end{aligned}$$

클래스내 스캐터(within-class scatter) 행렬

따라서 $\tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_w w$

LDA

마찬가지로 사영된 평균들의 간의 차이 (분산)를 원래의 특징공간에서의 평균들 간의 차이 (분산)으로 다음과 같이 동일한 형태로 표현된다.

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T\mu_1 - w^T\mu_2)^2$$

$$= w^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = w^T S_B w$$

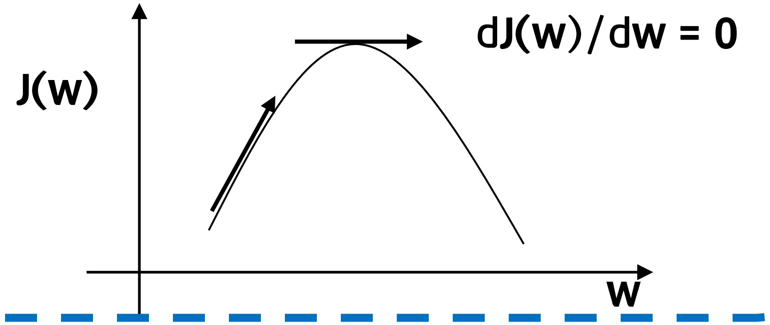
클래스간 스캐터(between-class scatter) 행렬

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{(\tilde{S}_1^2 + \tilde{S}_2^2)} \Rightarrow J(\mathbf{w}) = \frac{|\mathbf{w}^T\mathbf{S}_B\mathbf{w}|}{|\mathbf{w}^T\mathbf{S}_W\mathbf{w}|}$$

이걸 위해서 위에서 변환!

S_B 와 S_W로 표현된 Fisher의 기준 ➔ 이 목적함수를 최대로 하는 변환행렬 w 를 어떻게 찾을 것인가?

J(w)의 최대값을 찾기 위해서는 w 에 대하여 미분한 식을 0 으로 놓고, 이를 만족하는 w 를 구하면 된다.



Unit 01 | LDA

LDA

선형판별분석(LDA)에 의한 변환행렬 구성 및 분류 방법

1. 클래스간 분산(between-class scatter) S_B를 구한다.

$$\mathbf{S}_B = \sum_{i=1}^{c} N_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \qquad \boldsymbol{\mu} = \frac{1}{N}\sum_{\forall \mathbf{x}}^{c} \mathbf{x} = \frac{1}{N}\sum_{\mathbf{x}\in\omega} N_i\boldsymbol{\mu}_i \qquad \boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{\mathbf{x}\in\omega_i}\mathbf{x}$$

2. 각각의 클래스내 분산 (within-class scatter) S_{Wi}를 구한다.

$$\mathbf{S}_{W_i} = \sum_{\mathbf{x}\in\omega_i}(\mathbf{x}-\boldsymbol{\mu}_i)(\mathbf{x}-\boldsymbol{\mu}_i)^T$$

($\mathbf{S}_W = \sum_{i}^{C} \mathbf{S}_{W_i}$) ← 클래스-독립의 경우

3. 각각의 $\mathbf{S}_{W_i}^{-1}\mathbf{S}_B$ 에 대한 고유값 분석을 수행한다.

i클래스의 고유치들 중에서 q 개의 가장 큰 고유값 ($\lambda_1, \ldots, \lambda_q$)에 해당하는 고유벡터 ($u_1, \ldots, u_q$)를 선택하여 이를 열로 하는 i-클래스의 변환행렬 W_i 를 구성한다.

$$\mathbf{W}_i = [\mathbf{u}_1 \ldots \mathbf{u}_q]$$

4. 각각의 훈련자료들을 변환한다. $\mathbf{y} = \mathbf{W}_i^T\mathbf{x}$

5. 임의의 시험자료들을 변환하여(using W_i^T), 변환된 훈련자료의 평균과 비교하여 클래스를 결정

Unit 01 | LDA

LDA

이해 안가시는 분들은 갠톡주세요!