

[0207]_Dimensionality_Reduction_summary_SunjaeYoo

📅 날짜	@2023년 2월 7일
➤ 과목	
📁 구분	DSL
☑ 복습 1	<input type="checkbox"/>
☑ 복습 2	<input type="checkbox"/>
☑ 복습 3	<input type="checkbox"/>
📝 첨부	
📌 학기	etc

1 supervised vs unsupervised

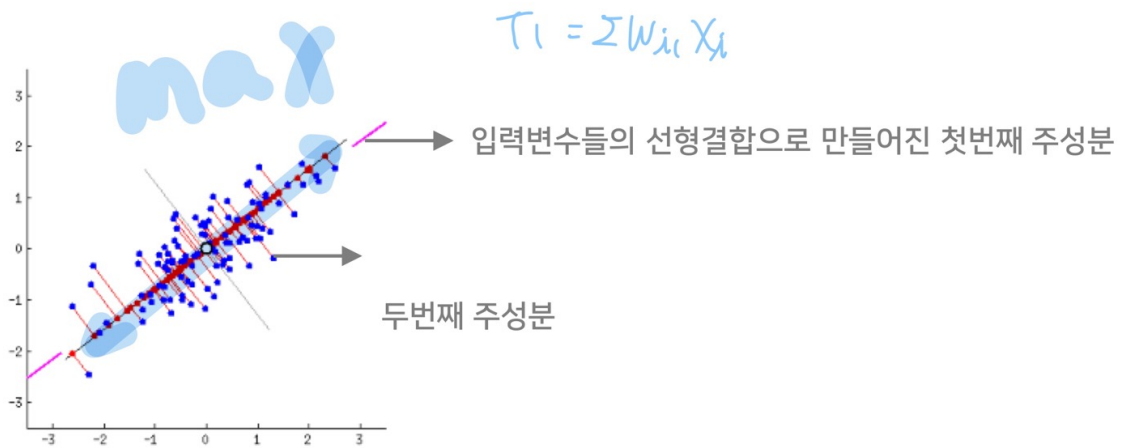
- supervised 는 y값을 고려하는 방법
 - Linear regression - 숫자
 - Logistic regression - 범주
 - SVM - 범주
 - Ensemble - 숫자, 범주

2 차원축소 - 특징선택 vs 특징 추출

- 특징 선택 : y값 고려 후 입력변수 선택 → supervised
 - mRMR : 입력변수 간 관계 최소화, 출력변수 간의 관계 최대화
 - $\arg\max$ (y와의 상관성 - 기존 입력변수들간 상관성의 평균) : forward
 - SVM-RFE : backward
 - Ridge, Lasso
- 특징 추출

		기존 입력변수 어떻게 조합할 것인가?	
		선형 결합	비선형 결합
Y라벨을 고려하여 특징을 생성할 것 인가?	Unsupervised	PCA MDS	KPCA Isomap LLE t-SNE AutoEncoder
	Supervised	LDA 병적	KFD

1) PCA - 잠재변수의 분산을 크게 만드는 로딩벡터를 구하고자 함. → 저차원에 mapping



분산 누적 → 80 ~ 90 % 분산을 설명하는 잠재변수 선택 후 잠재변수 만큼 차원 축소.

2) MDS - 원래 공간에서 데이터들 간 거리가 저차원 공간에서도 잘 보존되도록. → 저차원에 mapping

3) LDA - 클래스 구분이 잘 되도록 하는 특징 구하기.

4) KPCA - 커널함수 (sigmoid, poly 등) 을 이용해 고차원에서 PCA 진행

5) KFDA - 커널함수 (sigmoid, poly 등) 을 이용해 고차원에서 LDA 진행

6) Isomap - MDS의 확장판 : 모든 pairwise 거리를 잘 보존하도록 하는 저차원의 공간 구하기.

- 유클리디안 거리(MDS ; 저차원) 가 아닌 지오데식 거리!

7) LLE - 이웃들의 선형결합으로 잘 구축할 수 있는 가중치 벡터를 구하고, 이 가중치가 잘 보존되도록 하면서 저차원에 mapping.

- 낮은 차원에서도 최대한 유사한 가중치로 나와 이웃의 관계를 표현할 수 있도록!

8) t-SNE : 고차원 공간에서 비슷한 데이터 구조는 저차원 공간에서 가깝게 대응하고, 그렇지 않으면 멀리 떨어져 대응됨.

- 고차원의 확률분포와 저차원에서의 확률분포가 비슷하도록 !