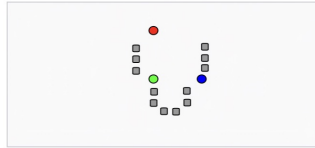


비지도학습: y값(레이블x). → 스스로 학습.

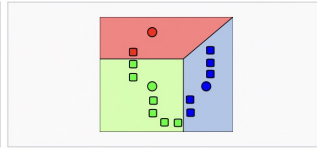
군집 (= Clustering)

표준 알고리즘의 실행 과정

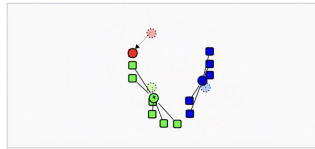
1. K 평균 분석



1) 초기 k "평균값" (위의 경우 $k=3$) 은 데이터 오브젝트 중에서 무작위로 뽑는다. (색칠된 동그라미로 표시됨).



2) k 각 데이터 오브젝트들은 가장 가까운 평균값을 기준으로 묶인다. 평균값을 기준으로 분할된 영역은 **보로노이 다이어그램**으로 표시된다..



3) k 개의 클러스터의 **중심점**을 기준으로 평균값이 재조정된다.



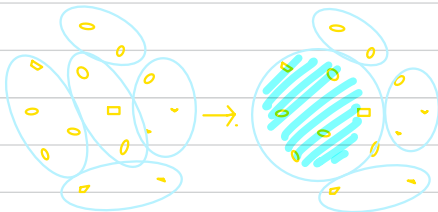
4) 수렴할 때까지 2), 3) 과정을 반복한다.

- **보로노이 다이어그램(Voronoi diagram)**은 평면을 특정 점까지의 거리가 가장 가까운 점의 집합으로 **분할**한 그림

- 각 군집내의 거리 차이 분산 **최소화**와 **방향**으로 동작.
- 레이블 없는 데이터에 레이블 달아줌 .

2.계층 군집 분석 (HCA)

- 가정: 거리가 가까운 **반응치**들은 **비슷한 특성**을 가질것임
- 두 클러스터 사이 거리 측정해 거리 가까운것끼리 묶는방식. → 계층적 방식 특징 또한 있음.



	목적	2층생성 방식.	군집 수
KNN	예측.	비계층적 (강렬).	자정.
HCA	데이터 축소.	계층적 (순차적).	자정 X.

순차적 분할 방법.

3. 기댓값 최대화 (EM 알고리즘).

가능도: 표본을 통해 맞을 찾는 방법. ^{ex)} 동전 던지기로 (H: H)가 나왔을 때 동전 확률이 $1/2$ 인 가설?

최대가능도: 표본 α 관측치 가장 가능한 값 θ 의 확률분포 찾는 것. (가능도 = α 인).

최대사후확률: 사후 확률의 최빈값.

최대가능도: 어떤 사건이 발생할 확률을 가장 높은 값 찾는 것.

최대사후확률: 최대우도 정리 + 맞은 사전확률과 결합된 확률로 구.

사전확률: 특정 사건 발생 이전 확률. (평균 μ 정음자).

관측되지 않는 잠재변수에 의존하는 확률 모델에서 최대가능도나 최대사후확률을 갖는 모수의 추정값을 찾는 반복적인 알고리즘

→ 최대우도 확률에 근접한 것

모수에 관한 추정값으로 로그가능도(log likelihood)의 기댓값을 계산하는 기댓값 (E) 단계와 이 기댓값을 최대화하는 모수 추정값들을 구하는 최대화 (M) 단계를 번갈아가면서 적용

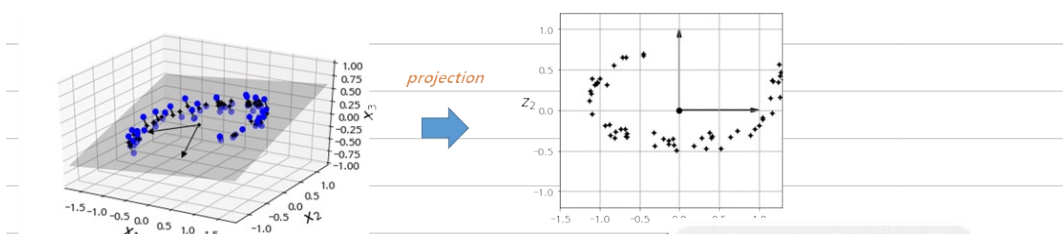
최대화 단계에서 계산한 변수값은 다음 기댓값 단계의 추정값으로 쓰인다.

시각화와 차원 축소

1. 주성분 분석

- 고차원의 데이터를 저차원의 데이터로 환원시키는 방법
- 서로 연관 가능성이 있는 고차원 공간의 표본들을 선형 연관성이 없는 저차원 공간(주성분)의 표본으로 변환
- 데이터를 한개의 축으로 사상시켰을 때 그 분산이 가장 커지는 축을 첫 번째 주성분, 두 번째로 커지는 축을 두 번째 주성분으로 놓이도록 새로운 좌표계로 데이터를 선형 변환
- 표본의 차이를 가장 잘 나타내는 성분들로 분해함으로써 데이터 분석에 여러가지 이점을 제공.

PCA는 데이터 하나 하나에 대한 성분을 분석하는 것이 아니라, 여러 데이터들이 모여 하나의 분포를 이룰 때 이 분포의 주 성분을 분석해 주는 방법



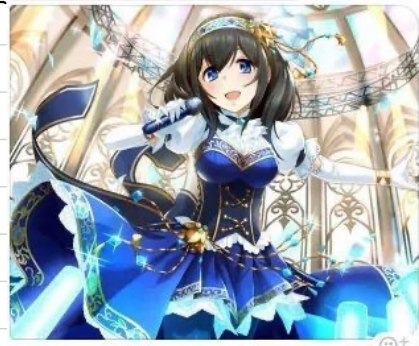
고차원에서는 데이터가 지평도
(3차원) : 4차원까지 가능.

저차원, 고차원

눈을 못 주면 뭐라도 생긴다길래 못했더니 정말 생겼습니다

2 커널 PCA

- PCA : 선형 타입.
- KPCA : 비선형 타입.



이쁘죠

오전 9:48

눈인거 아니가요?

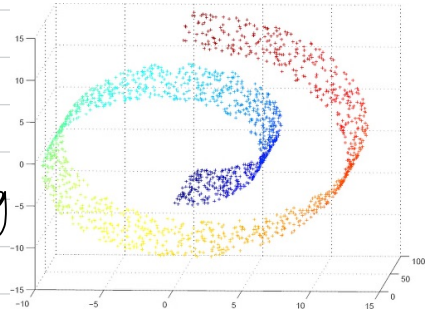
자원을 낮췄습니다

인정합니다

3. LLE(지형적 선형 임베딩)

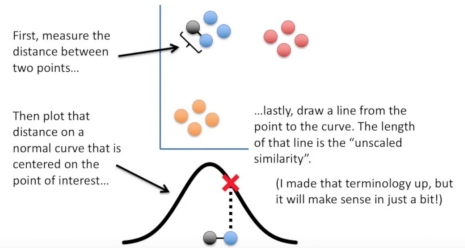
PCA: 차원 축소 하려다가 클래스끼리 섞여버림...

→ LLE는 고차원 공간에 입점한 데이터들의
구조 보존하며 저차원으로 embedding



4. t-SNE (PCA의 보완책).

차원 감소시 T분류에서 가까운 점들과 떨어진
편안해서 친밀도 높였기 때문



연관규칙 학습

1. 어프라이어리

특정 케이스 A가 빈발X면 A 포함하는 다른 케이스도 계산에서 제외시켜 계산속도↑

2. 이클릿

Apriori와 마찬가지로 빈발 항목 조합 마이닝 알고리즘

준지도학습

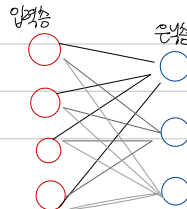
: 지도학습 + 비지도 학습.

DBN: 비지도 학습을 계층마다 진행. (이때 각 계층이 RBM).

RBM: 제한된 볼츠만 머신.

같은 층끼리는 연결X. ⇒ "제한"되었다.

DBN의 구성요소.



준지도학습 : 지도학습 + 비지도 학습.

DBN : 비지도 학습을 계층마다 진행. (이때 각 계층이 RBM).

RBM : 제한된 볼츠만 머신

같은 층끼리 연결 X \rightarrow '제한' 되었다.

DBN의 구성요소.

