복습

데이터셋의 값을 합치기도 한다. -> 데이터가 단순해진다. 처리 시간 단축

누락 데이터: 누락 데이터를 어떻게 처리할 지. 평균값, 최소, 최대 값을 임의로 넣을 것인지, 뺄것인지

적절한 알고리즘 선정

비지도 학습

지도 학습

강화 학습

회귀 분석

최적의 가중치는 어떻게 찾는가?

누가 더 중요한 예측자고 누가 덜 중요한 예측자인지

표준화도 가중치의 일종이다.

기울기 하강법(Gradient Descent)

두 개의 가중치가 있다면, 이왕이면 작은게 낫다

상관 계수

상관계수가 -1에서 1 사이에 나와야 하는데, 딱 -1이 나왔다. 자로 잰 것 같아서 사기 치는 게 아니냐고 의심받을 수 있다.

0이 나온다는 건 상관이 없다란 소리.

KNN

K개수가 너무 작으면 정확도는 높아지지만 수식이 너무 복잡해진다.

K개수가 크면 정확도가 떨어진다

적당한 정확도와 적당한 방정식으로 선을 긋는게 가능하다.

많이 실험해보는 수 밖에.

데이터 시각화에 유리

이런 경계선이 그려졌습니다---

지도 학습

서포트 벡터 머신

완충지대를 얼마나 결정할지가 중요하다

계산량이 상대적으로 적기 때문에 빠르다(knn보다)

RRRRRRRRRRRRRRRRRRRRRRRR