머신러닝(기계학습): 입력과 출력을 주어 규칙을 추출

AI: 입력과 규칙을 주어 결과를 예측

회귀문제(regression): mse 사용

분류문제(Classification): anccuracy 사용

차원축소

F(x)(규칙) = W(가중치)X(Input) + B(바이어스)

과소적합(Overfitting)

과대적합(Underfitting)

규제(Regulaization)

테스트와 검증

교차 검증: 데이터의 특성이 다르므로 데이터를 나눠서 훈련 데이터와 테스트 데이터를 바꿔가며 검증(홀드 아웃 검증) 후 평균을 구해 f(x)를 일반화 한다.

hitmap: feature들의 상관관계를 보기위해 사용

one-hot-encoding: 데이터가 가까이 있으면 상관관계가 생기는 현상을 끊기 위해

하이퍼파라미터 튜닝: 학습을 할 때 설정해주는 모든 값을 설정해주는 것

경사하강법: 함수의 값이 낮아지는 방향으로 각 독립변수들의 값을 변형시키면서 함수가 최솟값을 갖도록 하는 독립변수의 값을 탐색 방법. Weight와 bias 구하기

배치 경사하강법

확률적 경사하강법

홍일점 경사하강법.

Confusion matrix(혼동 행렬): 각 레이어 마다 뭐가 잘 분류 되고, 뭐가 잘 분류가 되지 않는지 구하기위해 사용

Accuracy: 정확도

F1(정밀도와 재현율의 조화평균): 불균형 데이터를 학습할 때 정확도만으로는 판단하기 힘들어서

Precision: 정밀도

Recall: 재현율

ROC곡선: 재현율과 1-특이도의 곡선을 그린 것

선형 회귀: x와 y의 상관관계를 구하는 것

로지스틱 회귀:

L1 Norm

L2 Norm

조기종료

Support Vector Machine

Decision Tree

Random Forest

지니 계수

지니 불순도

Bootstrapping

앙상블: 여러 개의 개별 모델을 조합하여 최적의 모델로 일반화하는 방법(ex. Random Forest)

Bootstrapping Aggregation(베깅)

앙상블 스태킹

차원의 저주

PCA(principal component analysis): 고차원 데이터를 저차원 데이터로 변환

샘플링 편향

스케일링 사용이유

리콜, 재현율, 정밀도

Logestic 회귀

소프트 백터

비지도학습

k평균

KNN: 지도학습, 분류

K-means: 비지도학습, 군집

퍼셉트론