

1 지난 시간에 소개한 내용들입니다.

컴퓨터시스템+머신러닝 => 새로운 능력

그 세세한 과정은 어떻게 될까요?

간단한 예제로 머신러닝을 활용하여 데이터로부터 대답을 얻는
세부 내용을 살펴봅시다.

2 한번 가정을 해볼게요.

이때 예측 시스템 => 모델

3 모델은 학습을 통해 완성되며, 모델의 목적은 정확한
예측입니다.

학습: 운전연습 예

지도학습/분류

돌발퀴즈~ 소맥이라면?

4 어떤 데이터를 수집?

두 가지

잡화점에서 구매

기타 수집 방법: 웹크롤링, open api, 설문, 센서측정 등

5 garbage in, garbage out

quality and quantity of data that you gather will directly
determine how good your predictive model can be.

측정: 결측치, 순서

테이블 형태로 만들어짐

=> 학습데이터

6 많은 시간 소요

1. 순서 랜덤: 순서 영향 제거
2. 결측치 처리
3. 인코딩! (feature engineering)

중복제거

정규화

오차 보정

Feature Engineering: feature engineering is the process of
using domain knowledge to extract features from raw data via
data mining techniques.

7 시각화: 데이터 이해 목적

- pandas/matplotlib
변수 간 상관관계 분석/통계

데이터 불균형: 치우쳐있는 데이터 <-> 현실세계와 => 오차요인

8 데이터 분리

- 대부분 Training
- 같으면 안됨 (math hw vs exam)
- Good train/eval split? 80/20, 70/30

9 다양한 모델이 있음: image, sequence, text, music

KNN, SVM, Decision Tree, Linear Model, Ridge, Lasso

10 two features

선형모델

11 알맞는 직선 찾기

12 처음엔 별로

나중엔 good

hyperparameters: learning rate => art, experimental, specifics
of dataset

13 테스트 데이터 이용

14 분류 모델 성능 평가 지표

F1 score는 정밀도와 재현율의 조화평균