

# 회귀 예측변수 (특성 feature) 종류에 따른 분류

지도 학습 알고리즘

- K-최근접 이웃 K-NN
- 선형 회귀
- 로지스틱 회귀
- 서포트 벡터 머신
- 결정 트리, 랜덤 포레스트
- 신경망

비지도 학습 알고리즘

- 군집
  - K-mean
  - DBSCAN
  - 이상치 탐지, 특이치 탐지
- 시각화, 차원 축소
  - 주성분 분석
- 연관 규칙

## 1. 선형 회귀

선형? 직선  $\Rightarrow$  일차 함수  $y = ax + b$

$y = 값$      $x =$  미리 주

회귀? 값 추정 (예측)

$a =$  기울기     $b =$  절편

선형 회귀 1차 함수로 표현  $\Rightarrow$  기울기 (a) 와 절편 (b) 찾기

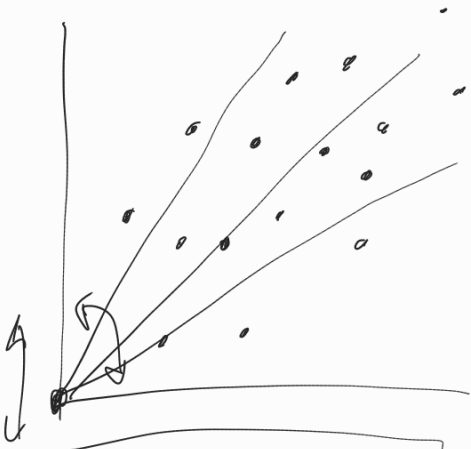
∴ 선형 회귀 알고리즘 = 기울기, 절편 찾는 알고리즘

X는 하나?  $\Rightarrow$  여러개이면?

X는 벡터 내적, 외적

↓  
해제

지도 학습		선형 회귀 : 1차식	
훈련 데이터		$y = ax + b$	
입력	레이블		
$x$	$y$		



Scikit - learn

- 1 colab
- 2 anaconda

## 2. 경사 하강법

$$y = ax + b$$

↑      단절점

기울기 사용해서 모델 계수씩 수정하는 최적화 알고리즘.  
 반복을  $\rightarrow$  미분.

$$\hat{y} = Wx + b$$

예측값      기울기      절편

(2차원  $y$ 와 관련하기 위함)

- 예측값으로 모델 출력  $(W, b)$

① 무작위로  $\{W, b\}$  선택

②  $x$ 에서 샘플 하나를 선택해서  $\hat{y}$  계산

③  $\hat{y}$  과  $y$  비교

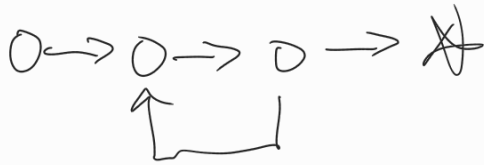
④  $\hat{y}$  과  $y$  가 가까워지도록  $w, b$  조절.

⑤ 모든 sample에 적용.

다 하려면 시간 얼마나?

최악의 경우  $n^2$  번.  $\rightarrow$  최적화 알고리즘 사용.

- 최적화 알고리즘



$\hat{y}$  과  $y$ 의 오차를 전달 (차이가 심하면  $(w, b)$  결정 어렵기 때문에)

✓ 예제 2: 전체 훈련 데이터 모두를 사용하여 한 번의 작업 진행

✓ 손실 함수:  $(y - \hat{y})^2$ 의 최솟값을 찾는다.

↳ 예측값, 실제값의 오차 계산

정리

선형 회귀  $\Rightarrow$  기울기, 절편

선형대수학

↓ 더 빠르게 계산

경사 하강법  $\Rightarrow$   $(y - \hat{y})^2$ 의 최솟값

미분

↓

확률적 경사 하강법  $\Rightarrow$  무작위로  $\{w, b\}$  선택

통계