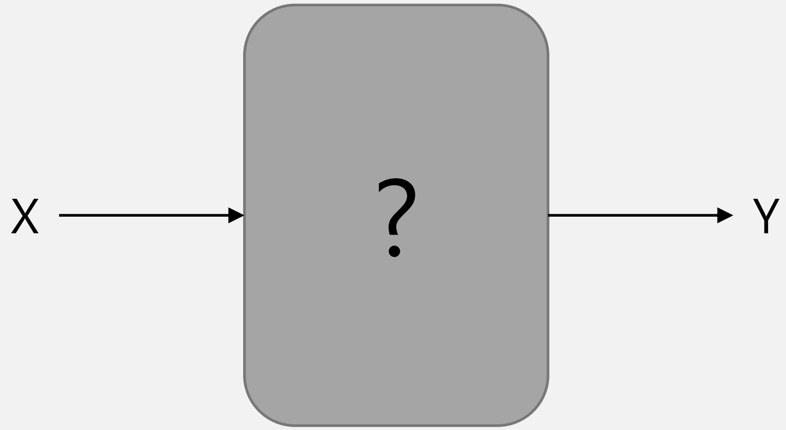
누구나 할 수 있는!

인터페이스 AI 오픈 스터디

2주차. 선형회귀와 지도학습

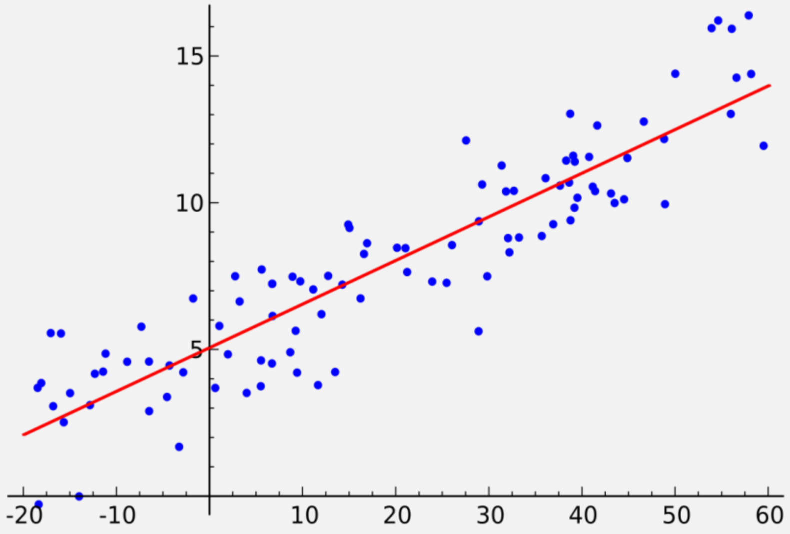
지도학습 복습

- 지도학습이란, 주어진 **입력에 대한 올바른 출력**을 기계에게 지도하는 것!
  - 그렇기 때문에, 입력과 출력(X,Y)의 데이터셋으로 기계를 학습



선형 회귀

- 선형 회귀란, 주어진 데이터의 경향성을 잘 나타내는 선형 관계를 찾아내는 것이다.
  - 일 평균기온이 x도 일 때, 아이스크림의 판매량 y는?
  - 연 소득이 x원 일 때, 매년 내야하는 소득세 y는?
  - 공부를 n시간 했을 때, 시험 성적 y는?



이미지출처

데이터 준비: 학습 시간에 따른 성적

- 아래와 같은 가상의 데이터를 만들어보자.

학습 시간	성적
2	60
4	70
7	85
10	100

학습 시간(X)와 성적(Y)의 관계를 어떤 식으로 나타낼 수 있을까?

가설 (Hypothesis)

- 학습 시간 X와 성적 Y가 아래 관계를 갖는다고 두자.

$$Y = WX + b$$

- W는 가중치, b는 편향(bias)이라고 부른다.
- 처음에는 랜덤한 값으로 W와 b를 초기화 한다.
- 우리는 각각 1, 1로 두고 시작해보자.

추측 (Prediction)

- 설정한 가설을 바탕으로 학습시간에 따른 성적을 예측해보자.

학습 시간	예측	성적
2	1*2+1 = 3	10
4	1*4+1 = 5	20
7	1*7 +1 = 8	35
10	1*10+1 = 11	50

W=1, b=1의 가설은 틀렸다! 〈- 당연한 결과

중요한 것은, 얼마나 틀렸는가?

비용 함수 (Cost Function)

- 비용 함수는 손실(Loss) 함수라고도 불린다.
- 우리가 답을 맞췄는지, 틀렸는지 (True or False)가 아니라, 얼마나 많이 틀렸는지 측정 가능한 기준을 제시해 준다!

예측	정답
9	10

예측	정답
-21	10

두 모델 중, 누가 더 나은 모델일까?

평균 제곱 오차 (Mean Squared Error)

- 평균 제곱 오차는 정답과 추측값 사이의 거리를 알려준다

$$MSE(T,Y) = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2$$

- 어려워 보인다면 직접 구해보자!

학습 시간	예측	성적	(t-y)	$(t-y)^2$
2	3	10	-7	49
4	5	20	-15	225
7	8	35	-27	729
10	11	50	-39	1521

MSE = 631

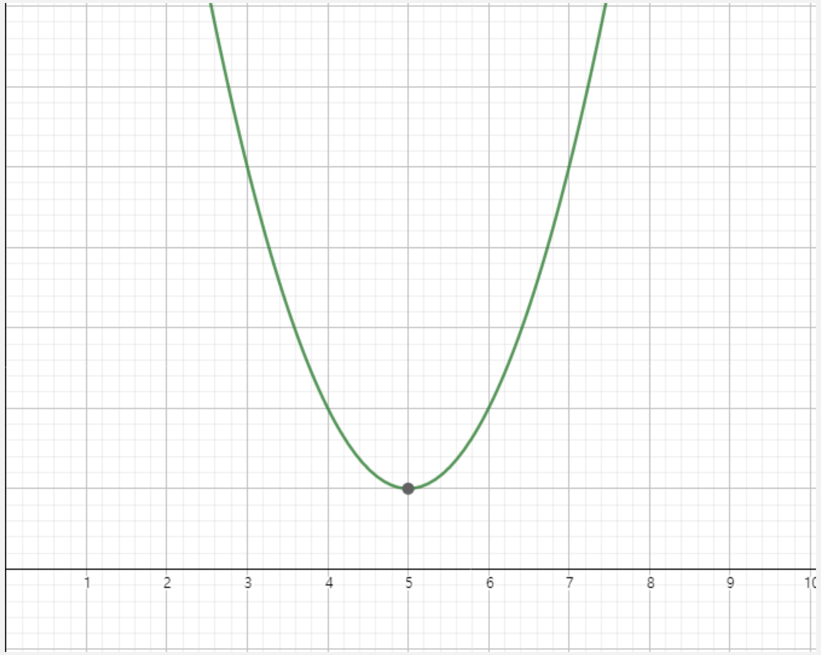
오차에 제곱을 해서 – 기호를 없에고 평균을 구하면..?

비용 함수의 조건

- 우리의 모델이 얼마나 틀렸는지(정답과 거리가 먼지)를 표현해야 함
- 이때, 이 값은 양수여야만 함
- 모델이 완벽히 정답과 일치하면, 비용 함수의 출력은 0
- 즉, 비용 함수를 0으로 만들면, 모델이 잘 학습된 것이다!
- 비용 함수 값이 작아질 수록, 모델이 정답에 가까워지는 것이다!

비용 함수의 그래프

- 성적 예측하기 문제의 비용 함수 그래프를, W에 대해 만들어보자.



W 값이 5에서, 비용 함수 값이 최소화된다.

다시 오차를 구해보면…

- W=5, b=1로 다시 추측해서 오차를 구해보자.

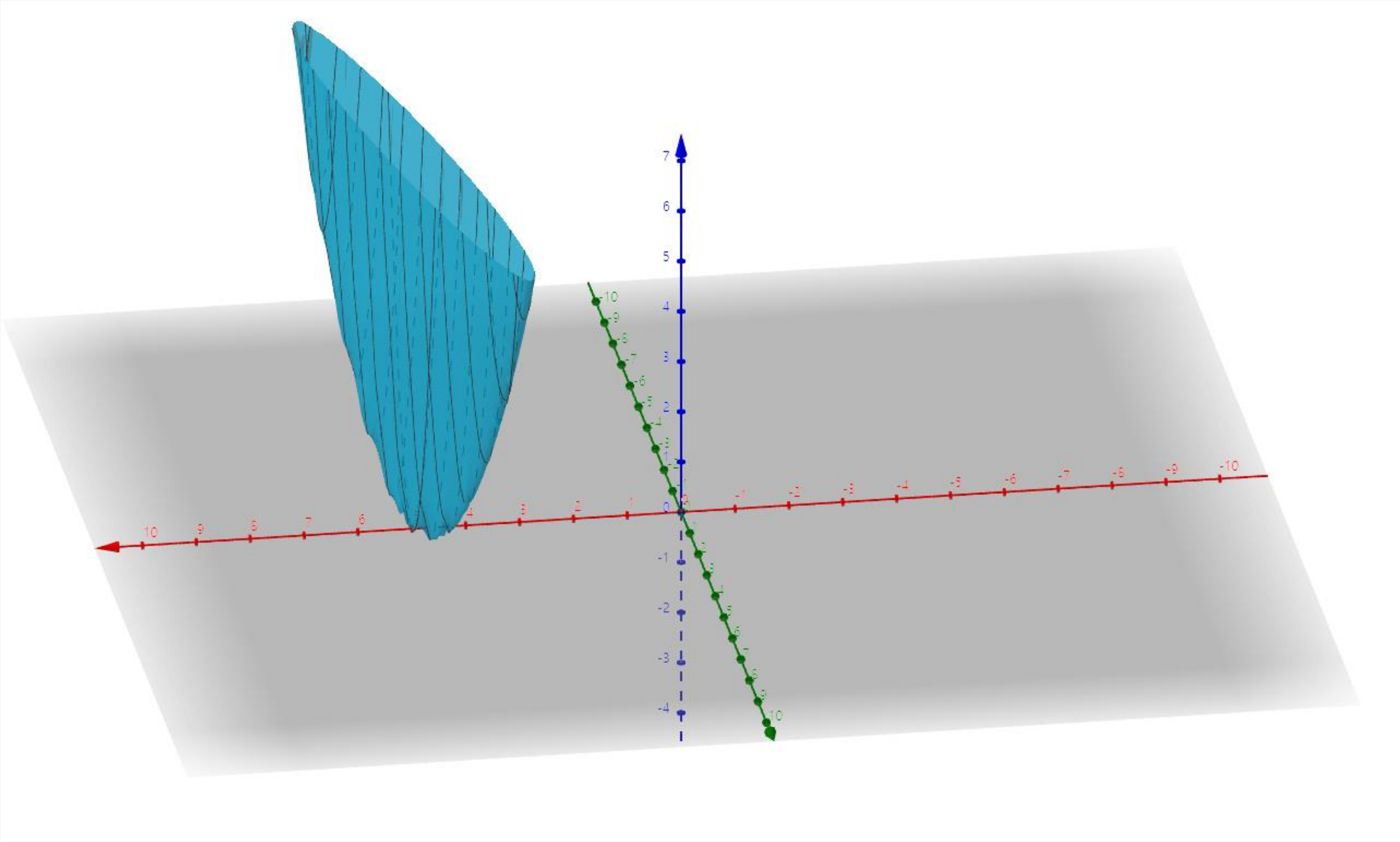
학습 시간	예측	성적	(t-y)	$(t-y)^2$
2	11	10	1	1
4	21	20	1	1
7	36	35	1	1
10	51	50	1	1

MSE = 1

B까지 0으로 최적화 해주면, 오차는 0이 된다!

비용 함수의 그래프 (b까지 고려)

- 최적화 해야 할 변수 W와 b에 대한 비용 함수의 그래프는 아래와 같다.



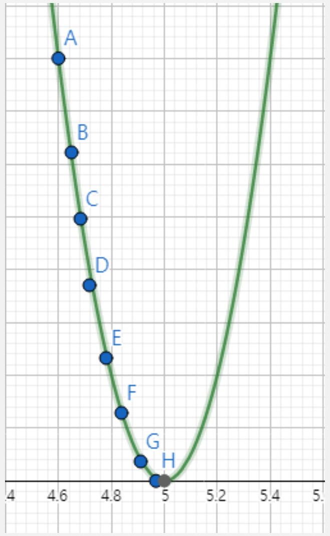
a(x,y)=0.33333 ((2x+y-10)^(2)+ (4x+y-20)^(2)+ (7x+y-35)^(2))

최적화를 어떻게 할 건데?

- 사람이야 그래프 그려보거나, 직관으로 W는 5, b는 0인게 눈에 보이지만, 컴퓨터는 아님.
- 비용 함수 값을 보고, 그 값을 최소화하는 W와 b를 찾아갈 방법이 필요함

경사 하강법 (Gradient Descent)

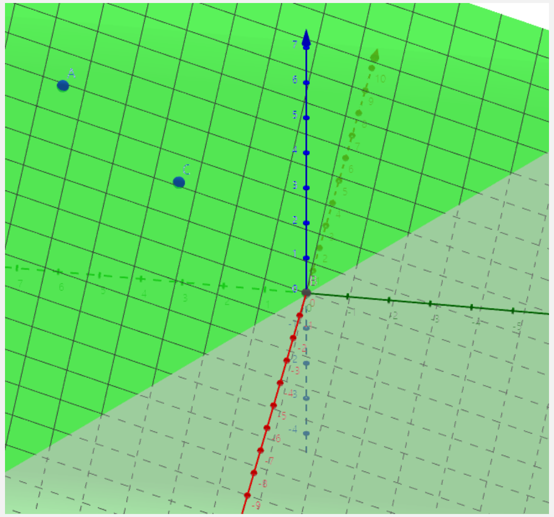
- 키워드
  - Gradient: 미분에서 기울기랑 같은 의미. 다만 2차원 이상에 많이 쓰임.
  - Descent: 하강, 높은 곳에서 낮은 곳으로 내려오다.



$$W_{n+1} = W_n - \gamma \nabla F W_n$$

W를 해에서 먼 초기점(A)에서 해 쪽으로 옮기기 위해, 경사(∇F)를 사용.
경사에 작은 값(γ)을 곱해 W에 조금씩 가감해서, W를 이동시킴.

변수가 2개 이상인 경우의 기울기의 표현



Z=0.2x+ 0.3y를 타고 내려오는 점

3차원 그래프에서, 기울기는 X에 대한 기울기 뿐 아니라 Y에 대한 기울기도 가짐.
그래서, 각 변수에 대한 기울기를 벡터 형태로 표현
Gradient = (0.2, 0.3)

경사를 구하는 방법: 편미분

- 각 변수에 대한 기울기를 따로 구해서 합치면 됨!
  - 0.2x + 0.3y 의 기울기는 x에 대한 기울기 0.2와 y에 대한 기울기 0.3을 합쳐서 (0.2, 0.3)의 벡터로 표현
- 식에서, 하나의 변수에 대한 기울기만 구하는 걸 편미분이라 함

$$F(w, b) = 2w + 3b$$

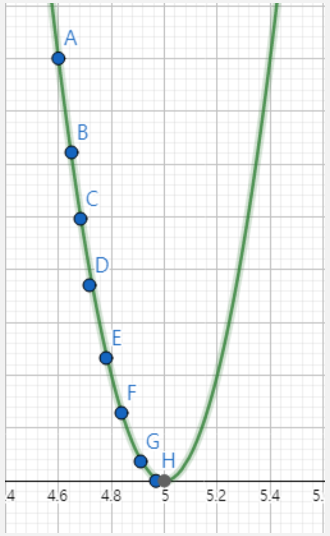
- 위 식에서, w에 대한 편미분을 하면 아래와 같음

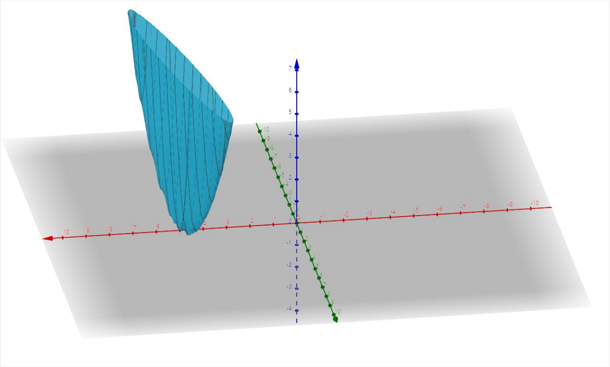
$$\frac{\partial F}{\partial w} = (2w)' = 2$$

- 위 식의 경사는 (2,3)

Convexity

- 경사 하강법으로 해를 잘 구하기 위해선, 문제가 Convex한 함수를 가져야 함.
- Convex한 함수는 하나의 해를 갖기에, 학습을 하다보면 적절한 해로 수렴함.

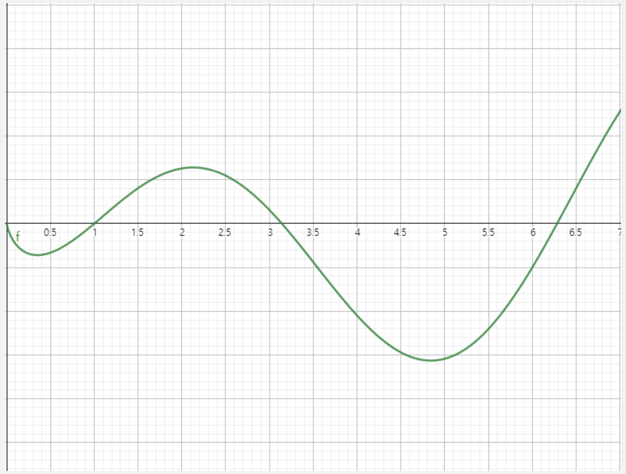




convex한 함수(지금까지 푼 선형 문제)

Convexity

- 대부분의 경우에는 non-convex 함.
- 그런데 또 생각보다 잘 됨. <- 현대 딥러닝의 큰 미스터리 중 하나
- Non-convex 한 함수는 해가 여러 개라서, 최고의 성능을 갖는 파라메터를 찾기 힘듦



Non-convex 함수

오늘의 키워드

- 지도학습 (Supervised Learning)
  - 가설 (Hypothesis)
  - 비용 함수 (Cost Function)
  - 경사 하강법 (Gradient Descent)
  - Convexity
- 선형 회귀 (Linear Regression)
  - 평균 제곱 오차 (Mean Squared Error)

감사합니다!