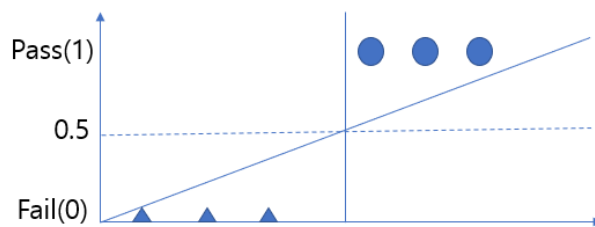


Logistic Regression : 이진 분류 문제일 때 사용

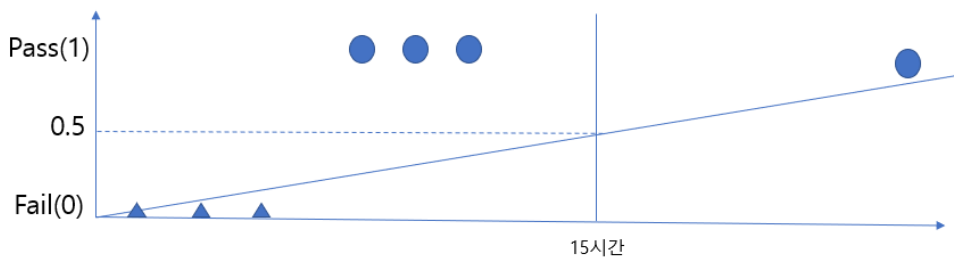
- sigmoid 함수, 새로운 cost 함수

< 선형 회귀로 분류 문제를 해결할 때의 문제점 >

예) 5시간 공부한 학생 - Pass (1)
 7시간 공부한 학생 - Pass (1)
 8시간 공부한 학생 - Pass (1)
 1시간 공부한 학생 - Fail (0)
 2시간 공부한 학생 - Fail (0)
 4시간 공부한 학생 - Fail (0)



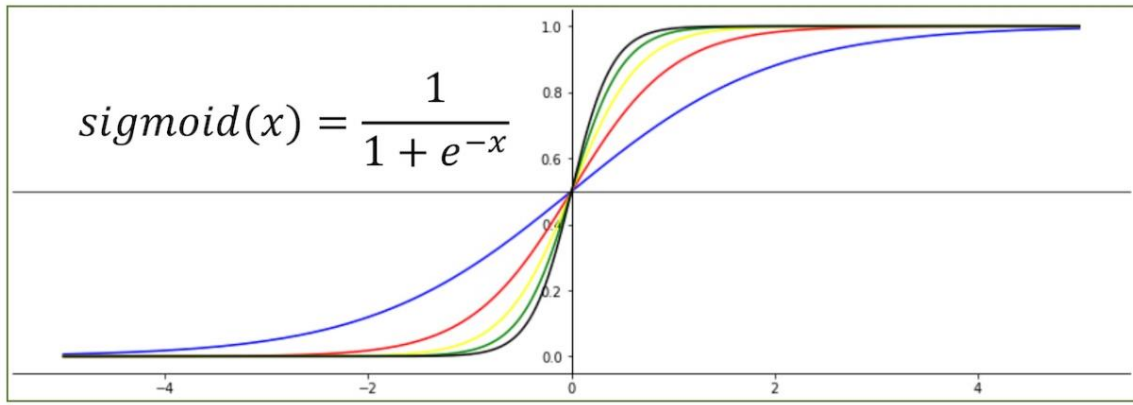
만약, 50 시간 공부한 학생이 있으면 ?



1) binary classification 의 결과는 0과 1이 필요한데,

$H(x) = Wx + b$ 는 1보다 훨씬 큰 수가 나타날 수 있음.

$H(x)$ 를 0~1 사이의 값으로 변경해주는 함수. \rightarrow sigmoid

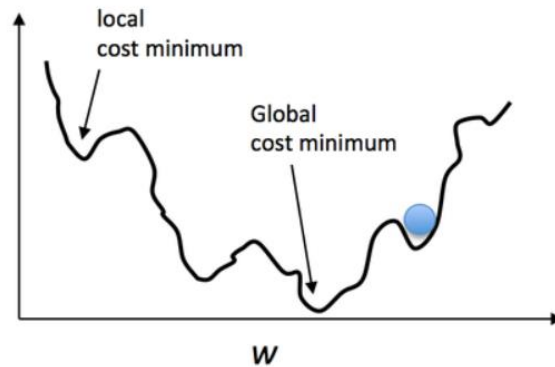


그래서, $H(x)$ 를 sigmoid 함수로 한번 더 계산. ($H(x)$ 를 0~1 사이의 값으로 변경)

$$\frac{1}{1 + e^{-H(x)}}$$

sigmoid 함수의 특징은, x 값이 아무리 크거나 작더라도 그 결과값은 최대가 1, 최소가 0이 됩니다.

2) sigmoid 함수를 통과한 값의 Cost 는 다음과 같이 구불구불한 모습으로 경사 하강법으로 최소 Cost 값을 찾을 때 문제가 발생할 수 있습니다. gradient descent algorithm 을 적용해도 우리가 필요로 하는 최종 값인 Global cost minimum 을 구하지 않고, local cost minimum 까지만 구할 수 있습니다.



이것을 해결하기 위해서 cost 함수를 아래와 같이 바꾸어야 합니다. 새로운 cost 함수

$$c(H(x), y) = \begin{cases} -\log(H(x)) & : y = 1 \\ -\log(1 - H(x)) & : y = 0 \end{cases}$$

회귀 분석과 달리 모수에 대한 t-검정, 신뢰 구간(confidence interval) 추정 등이 쉽지 않기 때문에 이를 보완하기 위해 다양한 성능 평가 기준이 필요하다

	Positive라고 예측	Negative라고 예측
실제 Positive	True Positive	False Negative
실제 Negative	False Positive	True Negative

참고) DS(Fraud Detection System)의 예

FDS(Fraud Detection System)는 금융 거래, 회계 장부 등에서 잘못된 거래, 사기 거래를 찾아내는 시스템을 말한다. FDS의 예측 결과가 Positive이면 사기 거래라고 예측한 것이고 Negative이면 정상 거래라고 예측한 것이다. 이 결과가 사실과 일치하는지 틀리는 지에 따라 다음과 같이 말한다.

True Positive: 사기를 사기라고 정확하게 예측

True Negative: 정상을 정상이라고 정확하게 예측

False Positive: 정상을 사기라고 잘못 예측

False Negative: 사기를 정상이라고 잘못 예측

Precision 정밀도

클래스에 속한다고 출력한 샘플 중 실제로 클래스에 속하는 샘플 수의 비율

FDS의 경우, 사기 거래라고 판단한 거래 중 실제 사기 거래의 비율. 유죄율

$$\text{precision} = \text{TP} / (\text{TP} + \text{FP})$$

Recall 재현율

TPR: true positive rate

실제 클래스에 속한 샘플 중에 클래스에 속한다고 출력한 샘플의 수

FDS의 경우, 실제 사기 거래 중에서 실제 사기 거래라고 예측한 거래의 비율.

검거율

$$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$$

F Score 정밀도(Precision)과 재현율(Recall)의 가중 조화 평균

$$\text{F Score} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$$

Accuracy 정확도 : $\text{accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$