Machine Learning

•Limitations of explicit programming

- Spam filter: many rules

- Automatic driving: too many rules

• Machine learning: "Field of study that gives

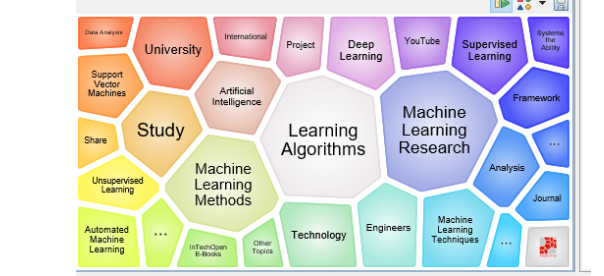
computers the ability to learn without being

explicitly programmed” Arthur Samuel (1959)



Supervised learning: - learning with labeled examples - training set

Unsupervised learning: un-labeled data - Google news grouping - Word clustering



Supervised learning

• Most common problem type in ML

- Image labeling: learning from tagged images

- Email spam filter: learning from labeled (spam or ham)

email

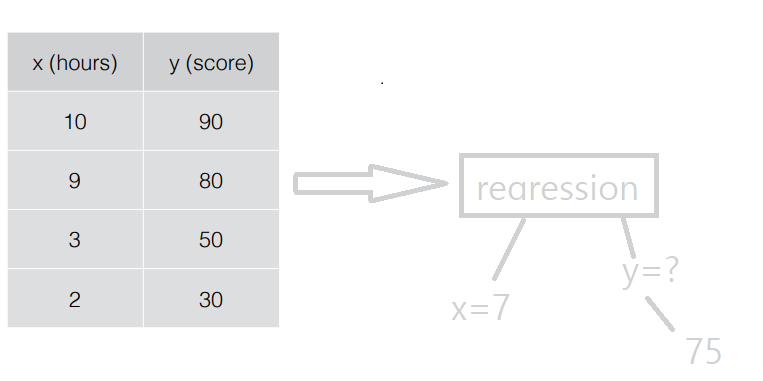
- Predicting exam score: learning from previous exam

score and time spent

Types of supervised learning

• Predicting final exam score based on time spent

- regression



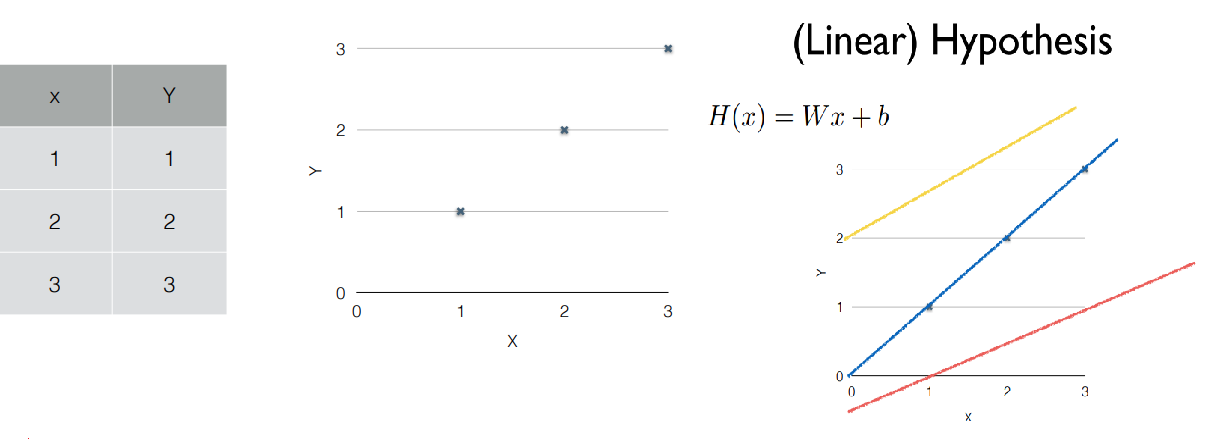
• Pass/non-pass based on time spent

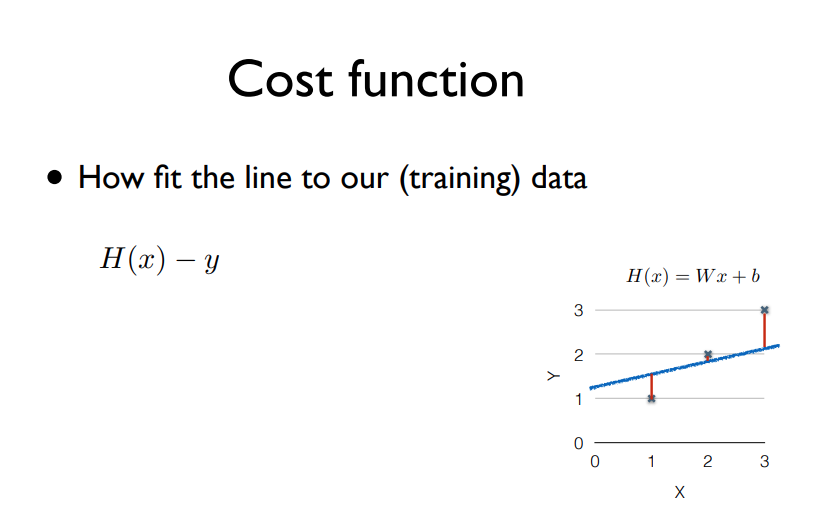
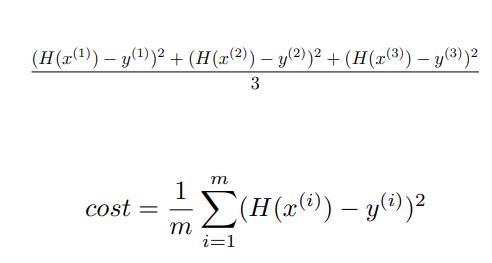
- binary classification

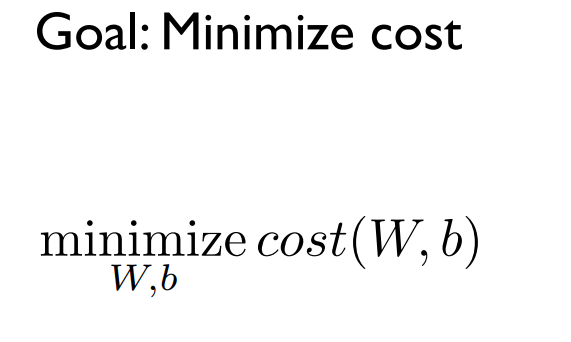
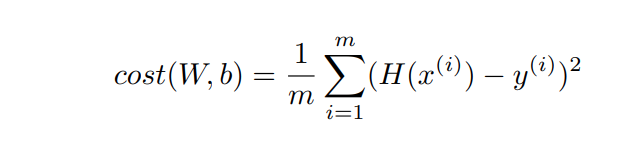
• Letter grade (A, B, C, E and F) based on time spent

- multi-label classification

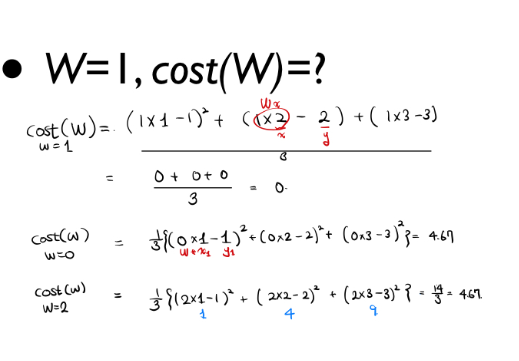
**Hypothesis**추론 모델을 의미한다. Linear Regression에서는 추론모델이 Linear 즉 다음과 같이 선형성이 있다고 가정을 하고 예측을 하는 것이다. 많은 경우에 Linear 모델로 설명할 수 있는 경우가 많기 때문에 자주 사용한다.





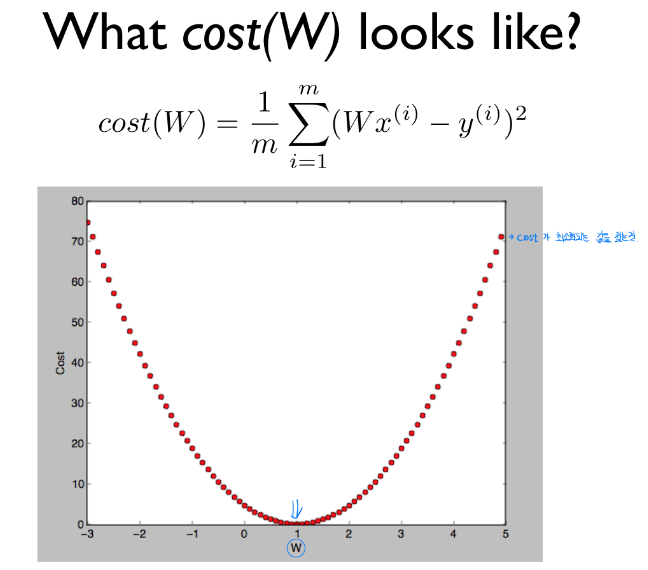




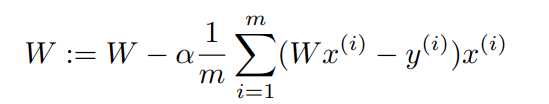


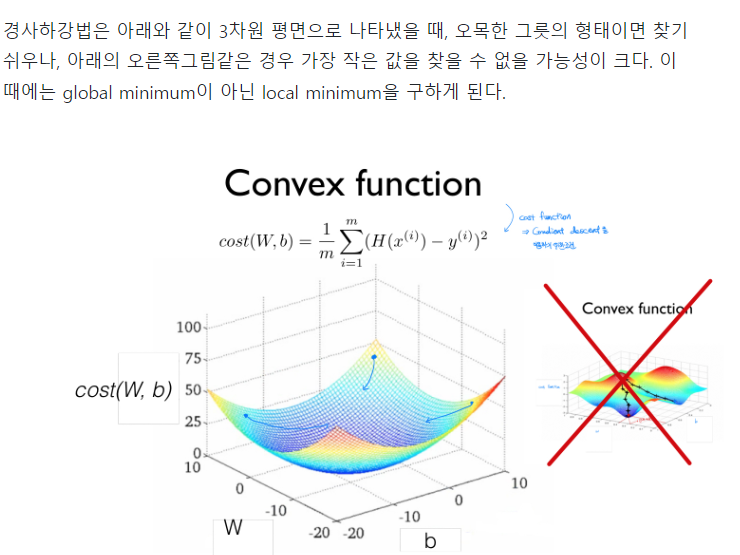
Gradient descent Algorithm(경사하강법)

• Minimize cost function • Gradient descent is used many minimization problems • For a given cost function, cost (W, b), it will find W, b to minimize cost • It can be applied to more general function: cost (w1, w2, …)



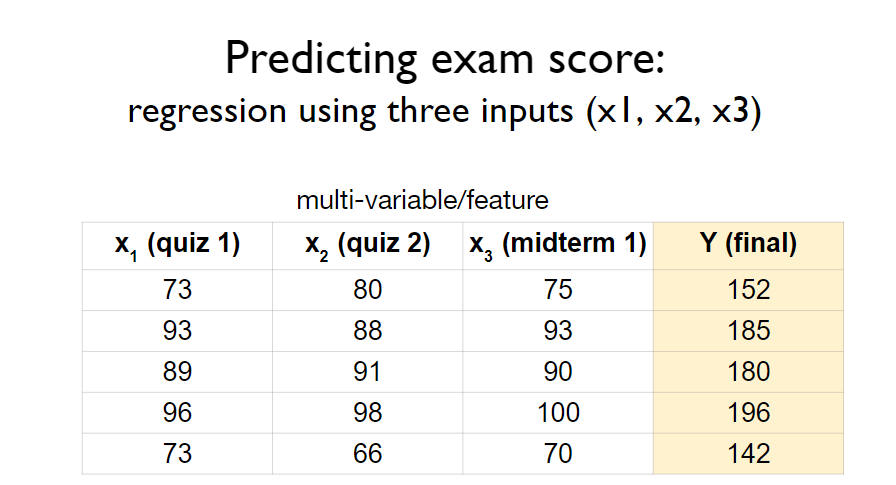
경사가 있는 쪽으로 내려가는 방식으로, 함숫 값의 미분을 통해서 경사를 찾고, learning rate와 미분 값의 곱만큼 내려가는 방식으로 찾아감

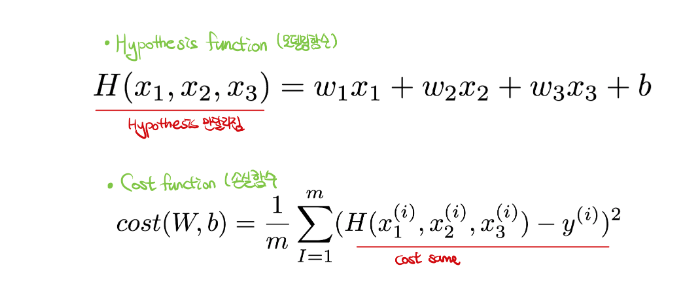




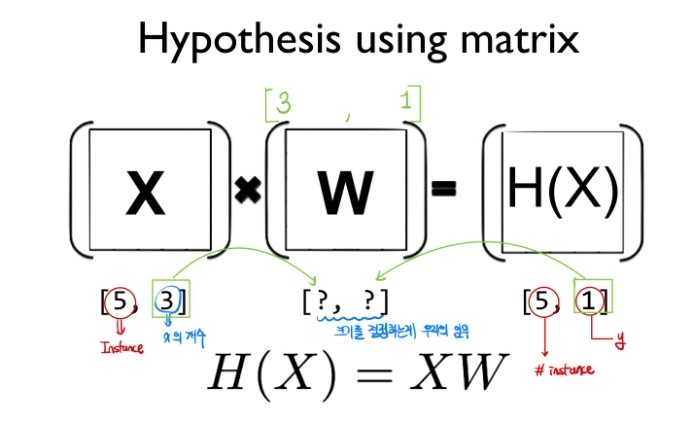
Local Minima로 빠지는 것을 방지하기 위한 방법으로 다양한 Optimize 기법있다.

Multi-Variable Regression





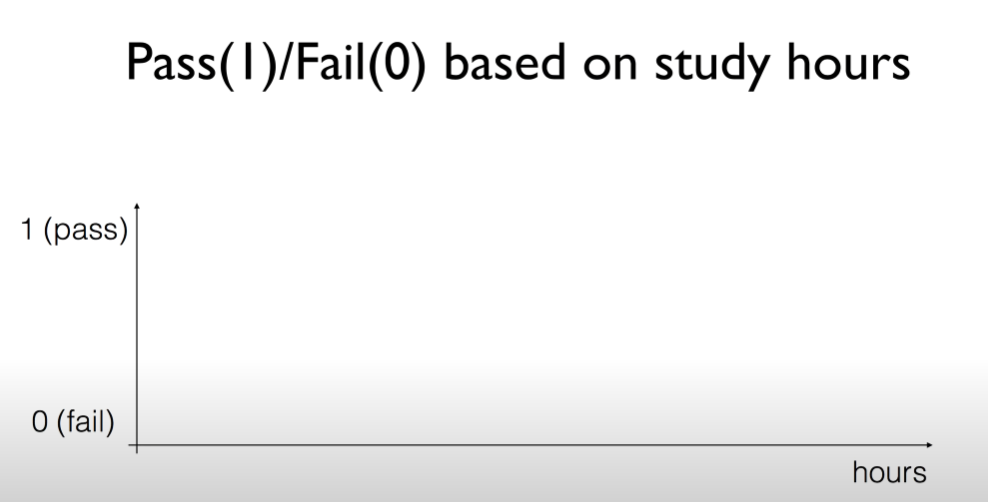
Multi-variable 함수에서는 계산의 효율성을 위해 다음과 같이 매트릭스를 이용해서 계산을 해준다.

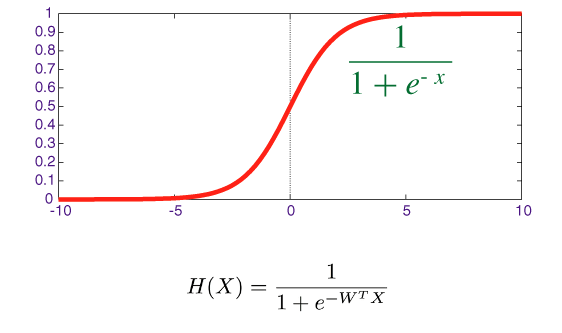


행렬쌍 계산을 할 때, X의 행의 갯수, 최종 행렬의 output의 숫자를 고려해서 W함수의 형태를 만들어준다.

Logistic (Regression) Classification

Classification 알고리즘들 중에서 굉장히 정확도가 높은 알고리즘으로 알려져 있다. 따라서 실제 문제에도 바로 적용해볼 수 있을 정도로 좋은 알고리즘이다





Logistic Hypothesis (가설 함수)

cost function을 이용하여 Logistic Regression

