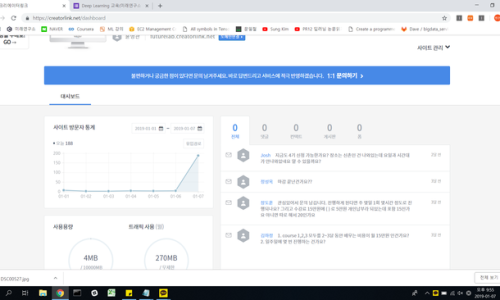
[http://futurelab.creatorlink.net/](http://futurelab.creatorlink.net/" \t "_blank)

[](http://futurelab.creatorlink.net/)

**[미래연구소](http://futurelab.creatorlink.net/" \t "_blank)**

[AI, 인공지능 Deep Learning beginner 미래연구소 딥러닝 입문 스터디 / 모집인원 : 25명 (선착순 마감) 수강료 : 월 15만원 / (Coursera 강의 수강료 월 5만원 개인결제)](http://futurelab.creatorlink.net/" \t "_blank)

[futurelab.creatorlink.net](http://futurelab.creatorlink.net/" \t "_blank)

오늘 배운 내용은 딥러닝이 머신러닝에 속하고, 머신러닝 역시 인공지능이라는 거대한 분야에 속한다는 것부터 시작해서, 머신러닝의 기초적인 것들까지 배웠다.

​

특히 머신러닝은 크게 두가지로 분류될 수 있는데, 이는 각 데이터값에 label을 붙여 학습을 시키기 때문에 '지도학습'이라는 뜻의 supervised learning과, label없이 데이터를 입력해 규칙을 찾아 결과값을 예측하는 unsupervised learning이 있다.

​

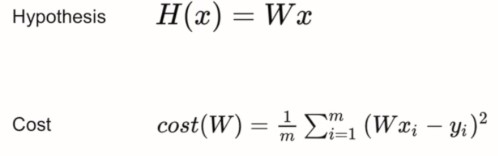
Supervised learning의 경우 결과값이 연속적인 regression, 결과값이 두가지인 경우의 binary classification, 그리고 결과값이 3개 이상인 경우의 multi-class classification이 있다. 따라서 예측하고자 하는 값이 무엇인지 그 특성에 따라 적용되는 기법이 달라진다.

​

예를 들어, 과거 시험성적을 토대로 앞으로의 시험성적을 예측하고 싶다고 하자. 받을 수 있는 성적은 A+부터 시작하여 F까지 여러 학점이 가능하므로, 이는 multi-class classification에 해당이 된다고 할 수 있다.

​

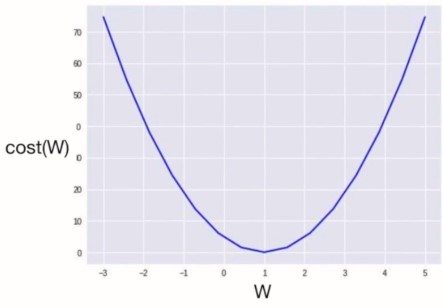
Linear Regression(선형회귀): 수많은 데이터가 주어지고 나서 y값이 도출되었을 때, 그 값들을 가장 잘 대변하는 직선은 y=ax+b(혹은 H(x) = Wx + b)으로 나타낼 수 있다. 이를 가설함수(hypothesis)라고 하는데, 이 가설함수를 통해 출력된 값들을 얼마나 잘 대변하고 있는지를 비용함수(cost function)를 통해 평가할 수 있다.

[](https://blog.naver.com/PostList.nhn?blogId=wordjun&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.naver.com%2Fmy.html&directAccess=true)

Hypothesis와 Cost함수의 산술적 정의

Cost(loss)는 hypothesis와 실제 데이터의 차이를 뜻한다. 그러나 어떤 부분에선 이 차이의 결과값이 양수일 수도, 음수일 수도 있으므로, 가설함수값에서 실제데이터 값을 뺀 것을 제곱해준다. 이렇게 모든 데이터의 차이점을 구한 값들의 평균값이 바로 cost가 된다.

​

[](https://blog.naver.com/PostList.nhn?blogId=wordjun&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.naver.com%2Fmy.html&directAccess=true)

딥러닝의 궁극적인 목표는 이 차이점을 최소화하는 것인데, 이는 경사하강법(gradient descent algorithm)을 통해 달성할 수 있다. 위의 cost함수를 통해 각 값들을 그래프에 나타내면 포물선이 그려진다. 머신러닝에서의 '학습'이란 것은 이 cost가 최저점이 되는 W값을 찾는 것이라고 할 수 있다.

따라서 위 그래프에서 cost값이 가장 작아지는 지점은 기울기가 0이고 W가 1인 지점이다.

​

인간은 위 그래프를 통해 한 눈에 어느 지점에서 cost값이 최소가 되는지 판단할 수 있지만, 컴퓨터는 최저점을 기계적으로 찾으려면 경사하강법을 사용해야 한다. 따라서 경사하강법은 cost를 최소화하는 알고리즘이라고 할 수 있다.

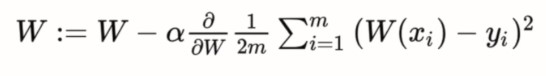
​

사용자가 정의한 시작점 혹은 임의의 점에서 시작하여, 해당 지점의 W값에 해당하는 기울기 값을 곱하여 W값에서 빼준다. 그렇게 되면 W는 아래로 조금씩 이동하게 되며(W값이 감소), 이 과정이 계속 반복되면서 결국 최저점에 달하게 된다.

​

기울기 값이 음수여도 같은 결과를 나타낸다. W값에 음수인 값을 곱하여 W에서 빼주게 되면, 그 결과는 양수이기 때문에 W값이 증가하게 된다. 따라서 W값은 증가하는 방향으로 이동하기 때문에, 기울기가 양수인 지점에서의 종착점과 동일한 결과를 보이게 된다.

​

[](https://blog.naver.com/PostList.nhn?blogId=wordjun&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.naver.com%2Fmy.html&directAccess=true)

각 지점에서의 기울기 값을 구하려면 2차함수인 cost함수를 미분해야 한다. 위 식에서 알파값은 learning rate를 뜻하는데, 이는 보통 매우 작은 값(0.00001)을 사용하며, W값에서 얼마나 뺄지를 결정하는 배수 역할을 맡는다. 만약 이 학습률이 크면 클수록 W에서 많은 값을 빠르게 빼며 큰 변화를 보일 것이고, 반대로 이 값이 작을수록 W값에서 더 작은 값을 빼며 느리게 학습하게 됨을 의미한다. 이 알파 값을 적당한 값으로 정하는 것이 매우 중요하다. 이 값이 너무 크다면 제대로 학습되지 않을 수 있고, 반대로 너무 작다면 학습 진행속도에 있어서 너무 느릴 수 있기 때문이다.