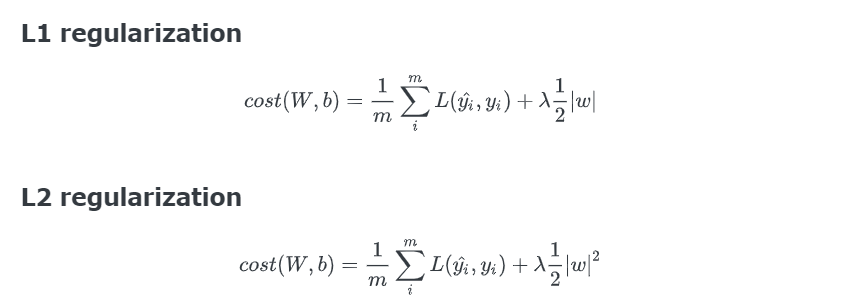
L1 Regularization, L2 Regularization의 비교

Regularization은 패널티를 부여하는 것!

왜? Overfitting을 막기위해서!



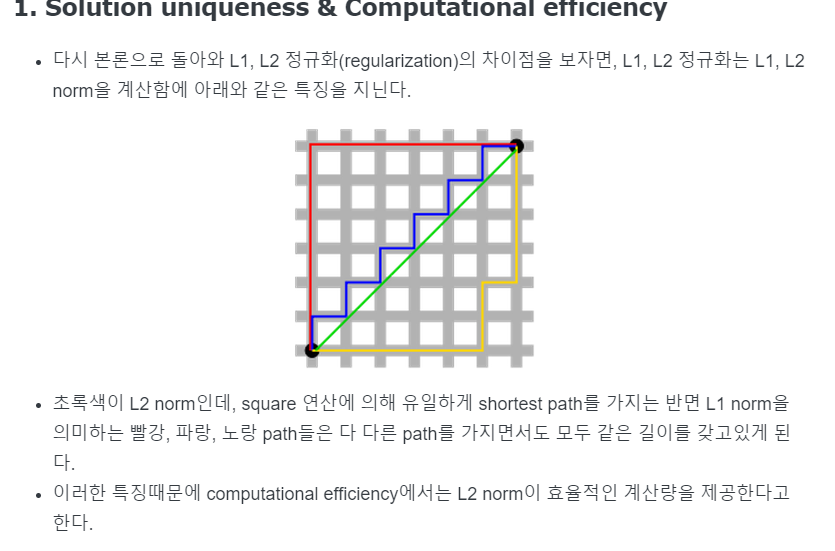
이 두 방식의 차이를 알기 위해선 Norm이라는 개념을 알아야 함

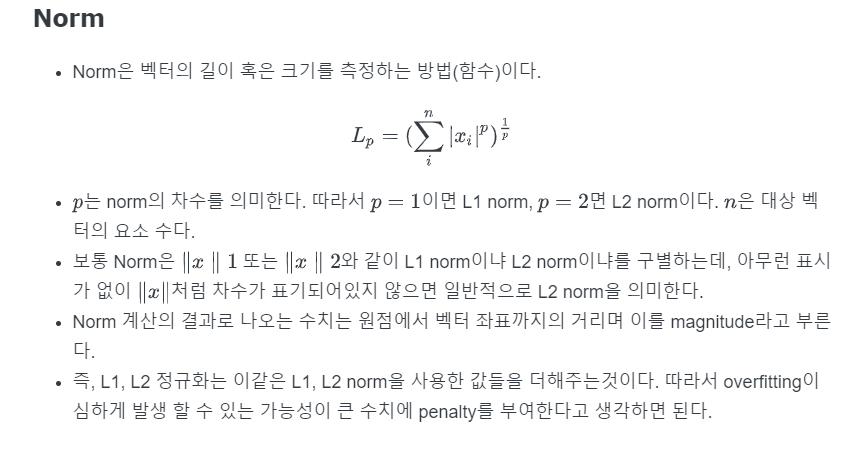
Norm은 벡터의 길이 혹은 크기를 측정하는 방법(함수)이다. (쉽게 생각해보자. 벡터는 1차원 배열이야. A = {1, 2, 3, 4}가 있다고 할 때 A의 Norm은 1+2+3+4 해서 7이 될 수도 있고(L1 norm)

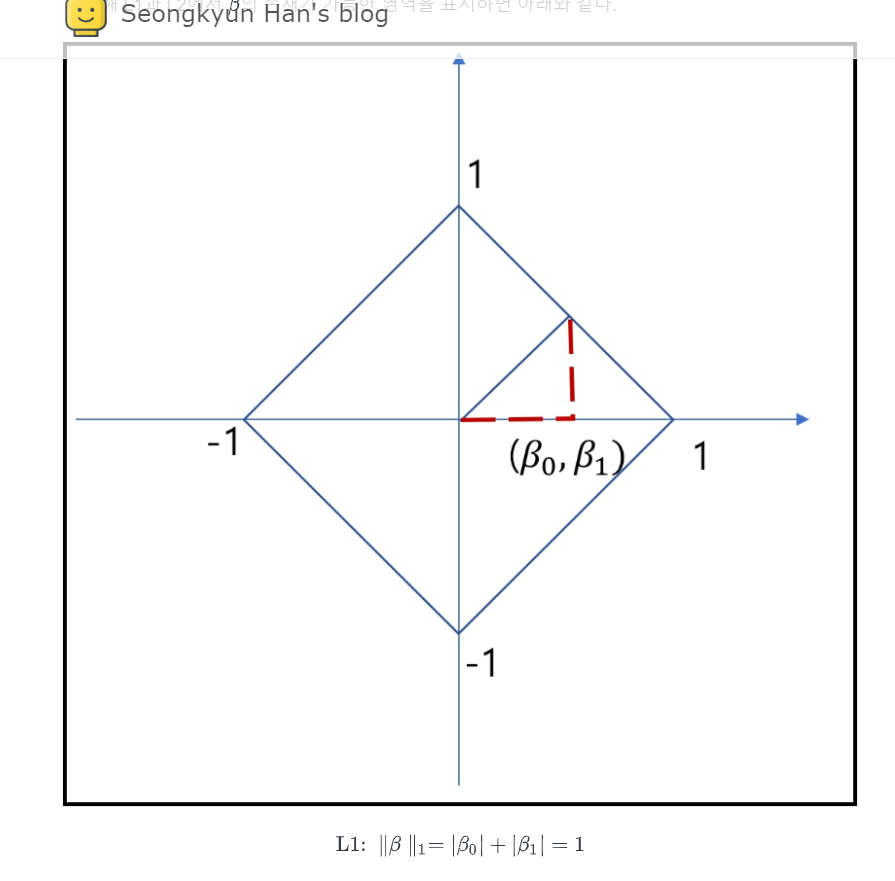
루트(1^2 + 2^2 + 3^2 + 4^2)해서 루트30이 될수도 있음(L2 norm))

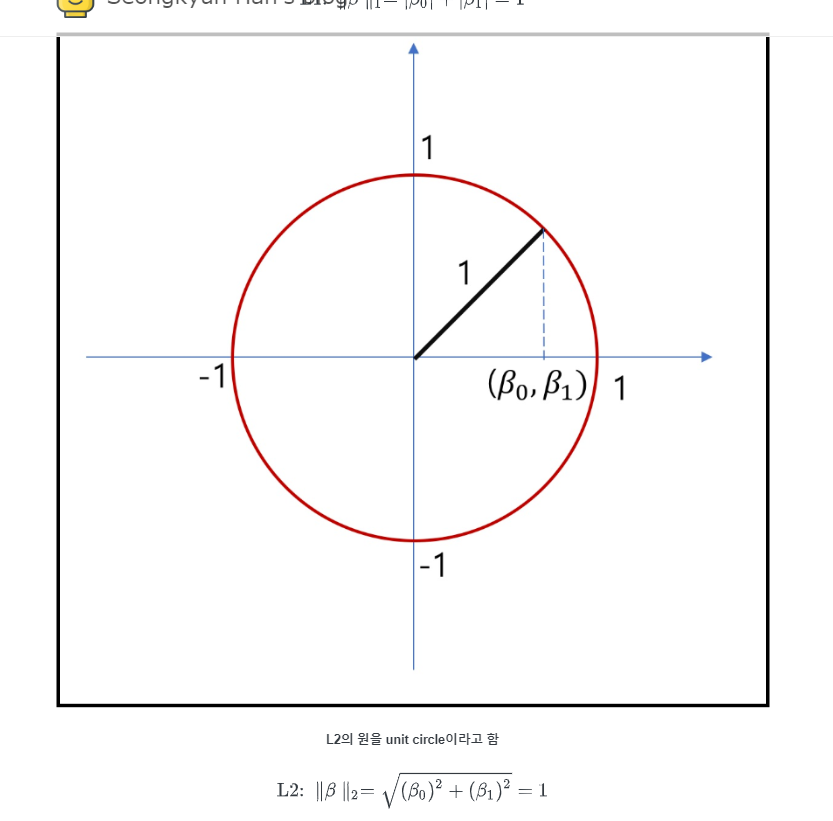
이게 아래 그림에 대한 이해가 될 수 있음.

빨간선, 노란선, 파란선은 모두 거리 7을 나타내는 L1 norm인 것이고 초록색 선은 거리 루트30을 나타내는 L2 norm인 것임.







위 그림은 L1 정규화에 대한 설명. B0, B1이 각각 x, y 라고 생각하면 됨. 

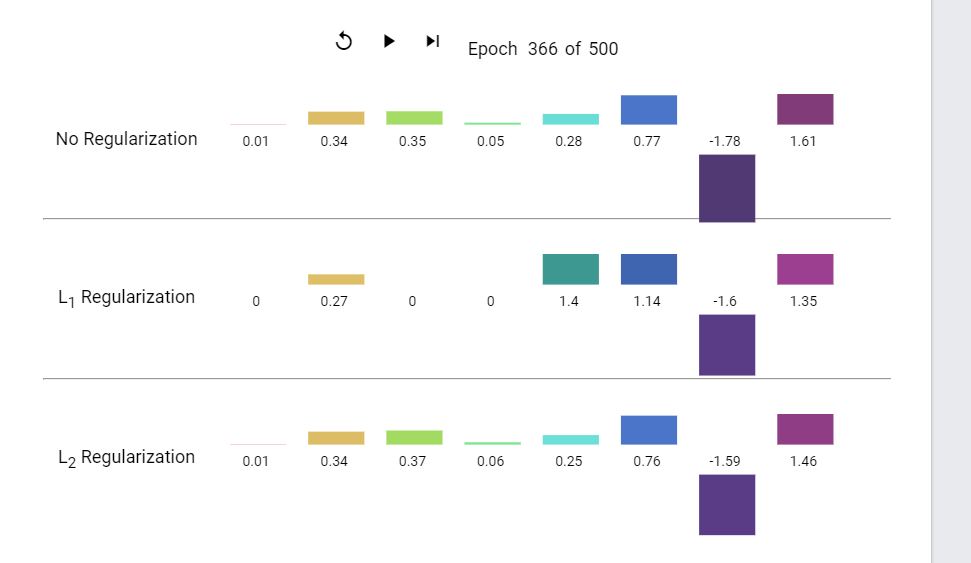
L2 norm

L1은 feature selection이 가능하다고 한다. Feature란 벡터의 요소를 말하는 것

쉽게 생각해보면 1+2+3+4 = 10인 것은 1 + 2\*0 + 5 + 4 = 10으로 대체 가능하다.

이런식으로 많은 coefficient(계수)들을 0으로 만들면 sparse한 형태를 취할 수 있다!. Sparse matrix는 거의 대부분의 행렬값들이 0인 행렬을 뜻한다.

이걸 뭉뚱그려서 생각해보면 같은 값을 내는데 연산을 줄일 수 있을 것을 보인다!



출처 : <https://developers.google.com/machine-learning/crash-course/regularization-for-sparsity/l1-regularization>

이거 실행해보면 에폭이 커질수록 L1 Regularization을 하는 weight value들의 값이 0으로 떨어지는게 몇 개 보이는 걸 알 수 있다.