L2 – Regularization / Ridge

Linear Regression

Director of TEAMLAB
Sungchul Choi

L2 regularization

- 기존 Cost function L2(norm) penalty term을 추가

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$

- norm - 벡터의 길이 혹은 크기를 측정하는 방법

$\|(\theta)\|_2^2$ L2는 Euclidean distance
원점에서 벡터 좌표까지의 거리

L2 regularization

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\theta_0 := \theta_0 - \alpha \ \frac{1}{m} \ \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \ \left[\left(\frac{1}{m} \ \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right] j \in \{1, 2 ... n\} \qquad \}$$

L2 regularization

$$\theta_j := \theta_j - \alpha \left[\left(\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right]$$

$$\theta_j := \theta_j (1 - \alpha \frac{\lambda}{m}) - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Normal equation approach

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\begin{aligned}
J(\theta) &= (y - X\theta)^T (y - X\theta) + \lambda\theta^T\theta \\
&= y^T y - \theta^T X^T y - y^T X\theta + \theta^T X^T X\theta + \lambda\theta^T\theta \\
&= y^T y - \theta^T X^T y - \theta^T X^T y + \theta^T X^T X\theta + \theta^T \lambda I\theta \\
&= y^T y - 2\theta^T X^T y + \theta^T (X^T X + \lambda I)\theta
\end{aligned}$$

Normal equation approach

$$J(\theta) = y^T y - 2\theta^T X^T y + \theta^T (X^T X + \lambda I)\theta$$

$$\frac{\partial J(\theta)}{\partial \theta} = -2X^T y + 2(X^T X + \lambda I)\theta$$

$$(X^T X + \lambda I)\theta = X^T y \rightarrow \hat{\theta} = (X^T X + \lambda I)^{-1} X^T y$$

L1 regularization

- 기존 Cost function L1(norm) penalty term을 추가

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n}|\theta_j|$$

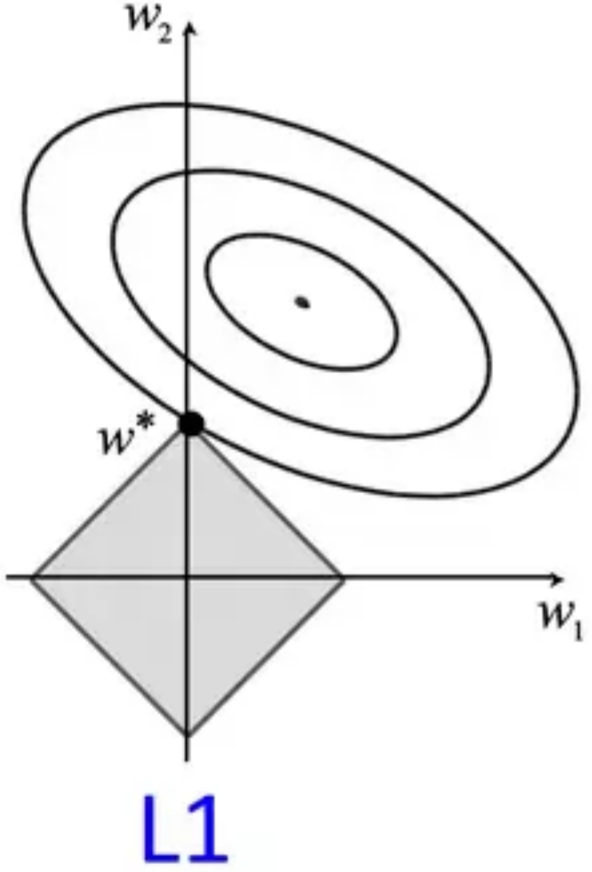
- norm - 벡터의 길이 혹은 크기를 측정하는 방법

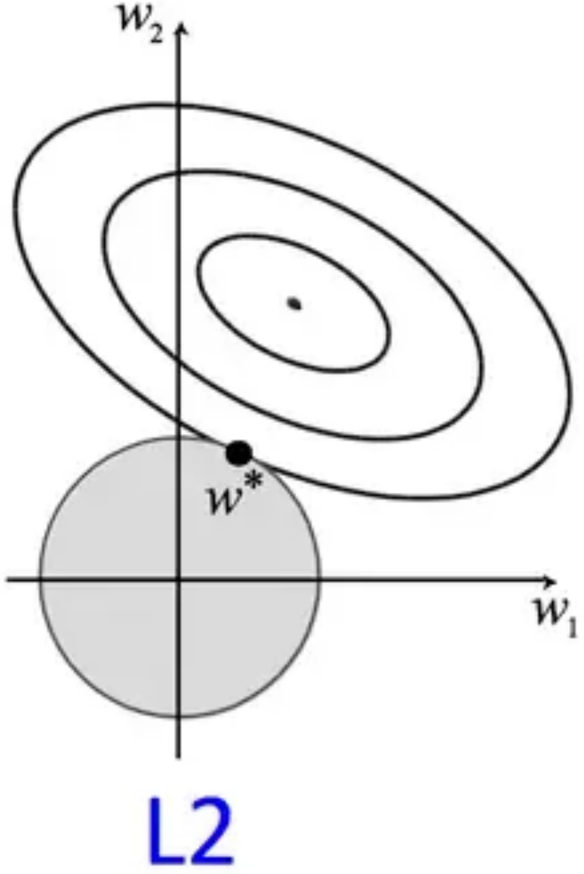
$\|x\|_1 := \sum_{i=1}^{n} |x_i|$ L1는 manhattan distance
원점에서 벡터 좌표까지의 거리

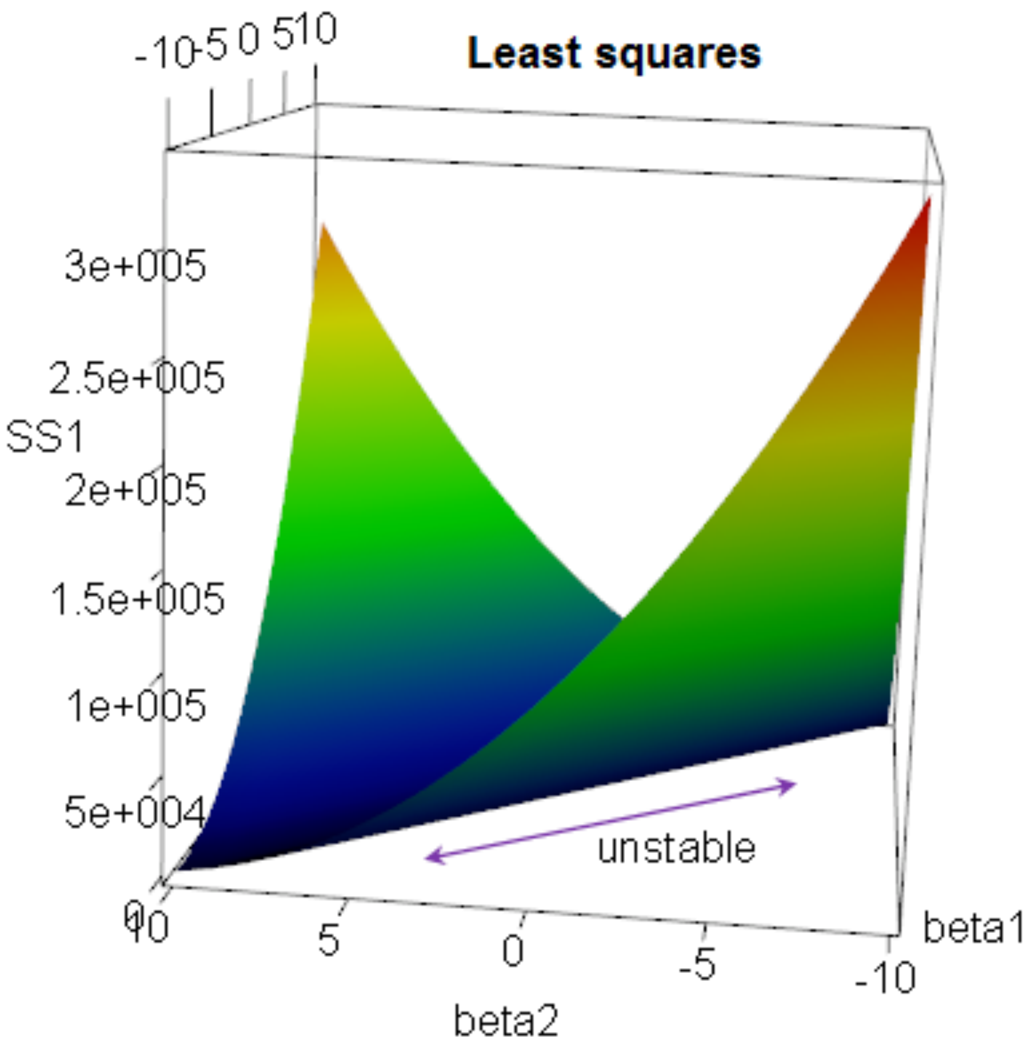
http://taewan.kim/post/norm/

L1 vs L2

$$\sum_{j=1}^{2} |w_i| \leq s \qquad\qquad \sum_{j=1}^{2} (w_i)^2 \leq s$$







https://stats.stackexchange.com/questions/151304/why-is-ridge-regression-called-ridge-why-is-it-needed-and-what-happens-when

L1	L2
Unstable solution	Stable solution
Always on solution	Only one solution
Sparse solution	Non-sparse solution
Feature selection	

TEAMLAB

Human knowledge belongs to the world.