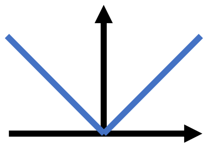
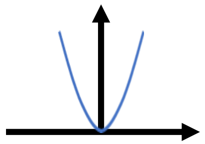
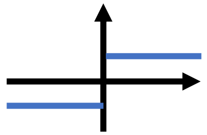
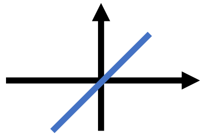
Lasso vs. Ridge

구분	Lasso	Ridge
수식	$\|\boldsymbol{w}\|_1$ 	$\frac{1}{2}\|\boldsymbol{w}\|^2$ 
미분	$\begin{cases} 1 & w < 0 \\ -1 & w > 0 \end{cases}$ 	$\boldsymbol{w}$ 
특성	• 가중치 값을 정확하게 0으로 만듦 • 중요한 특징을 ‘선택’하는 효과 • 모델에 Sparsity를 가함.	• 큰 가중치의 값을 작게 만듦 • 모델 전반적인 복잡도를 감소시키는 효과 • 가중치의 값이 0이 되게 하지는 못함

L-1 Loss, L2-Loss

$$Cost = Loss(Data|Model) + \lambda Complexity(Model)$$

$$y = wx$$

$$\arg\min_{w} \sum \|y_i - wx_i\|_2^2$$

MSE Loss

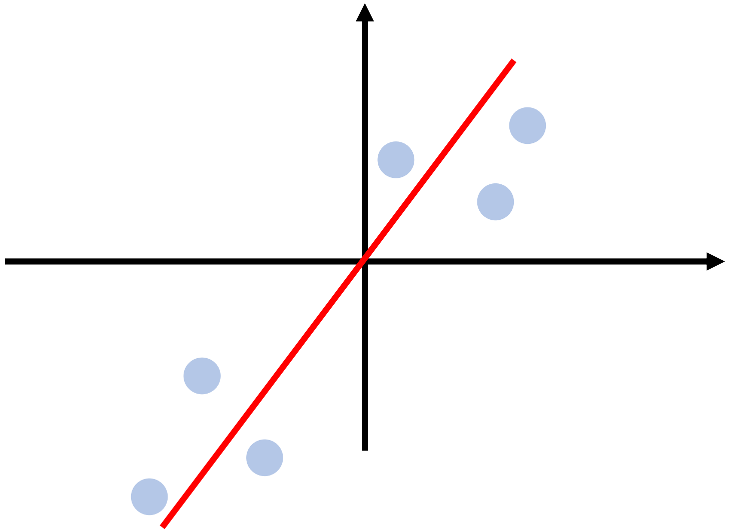
$$\arg\min_{w} \sum |y_i - wx_i|$$

MAE Loss

정규화를 통해서 L-1 Norm과 L-2 Norm의 특성에 대해 살펴보았다.

이 기회에 Linear Regression에서 L-1 Loss와 L-2 Loss의 특성까지 조금만 더 살펴보자.

MSE Loss

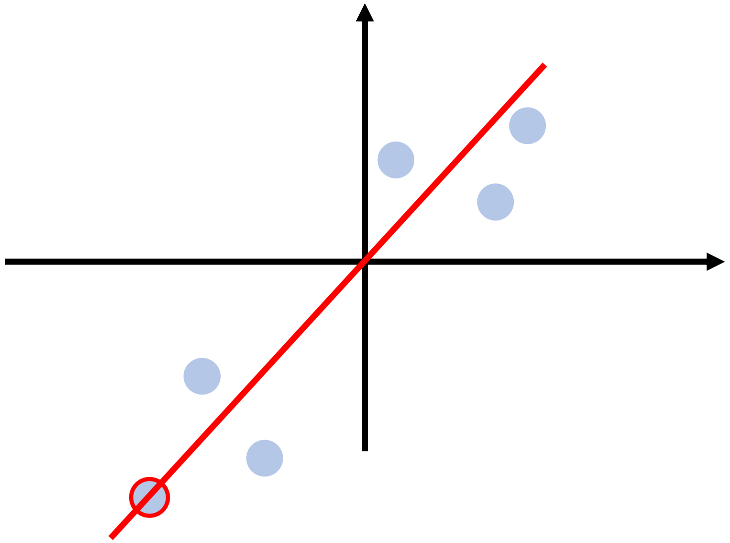


$$\arg\min_{w} \sum \|y_i - wx_i\|_2^2$$

- 에러가 클 수록 더욱 큰 페널티
- 데이터의 평균 – 존재하지 않는 값
- 데이터를 Smoothing하는 효과
- Outlier에 취약한 단점이 있음

보통 별 의심 없이 써 오던 MSE Loss는 '평균'을 나타내는 특성이 있다.

MSE Loss



$$\arg\min_{w} \sum |y_i - wx_i|$$

- 에러가 커져도 동일한 페널티
- 데이터의 중간값 – 존재하는 정확한 값
- 존재하는 값을 사용하여 샤프한 특성
- 적게 존재하는 값을 무시하는 특성

MAE Loss는 ‘중간값’의 특성이 있으며, Outlier에 강건한 특성이 있다.