회귀 시에 사용하는 손실 함수 탐구

Introduction

· 모두 Regression 태스크에서 자주 사용되는 loss functions이다.

· 이들의 차이는 무엇이고, 언제 사용하는 것이 적합할까?

Mean Absolute Error

$$MAE = \frac{1}{N}\sum_{i}^{N} |(pred_i - target_i)|$$

Mean Squared Error

$$MSE = \frac{1}{N}\sum_{i}^{N} (pred_i - target_i)^2$$

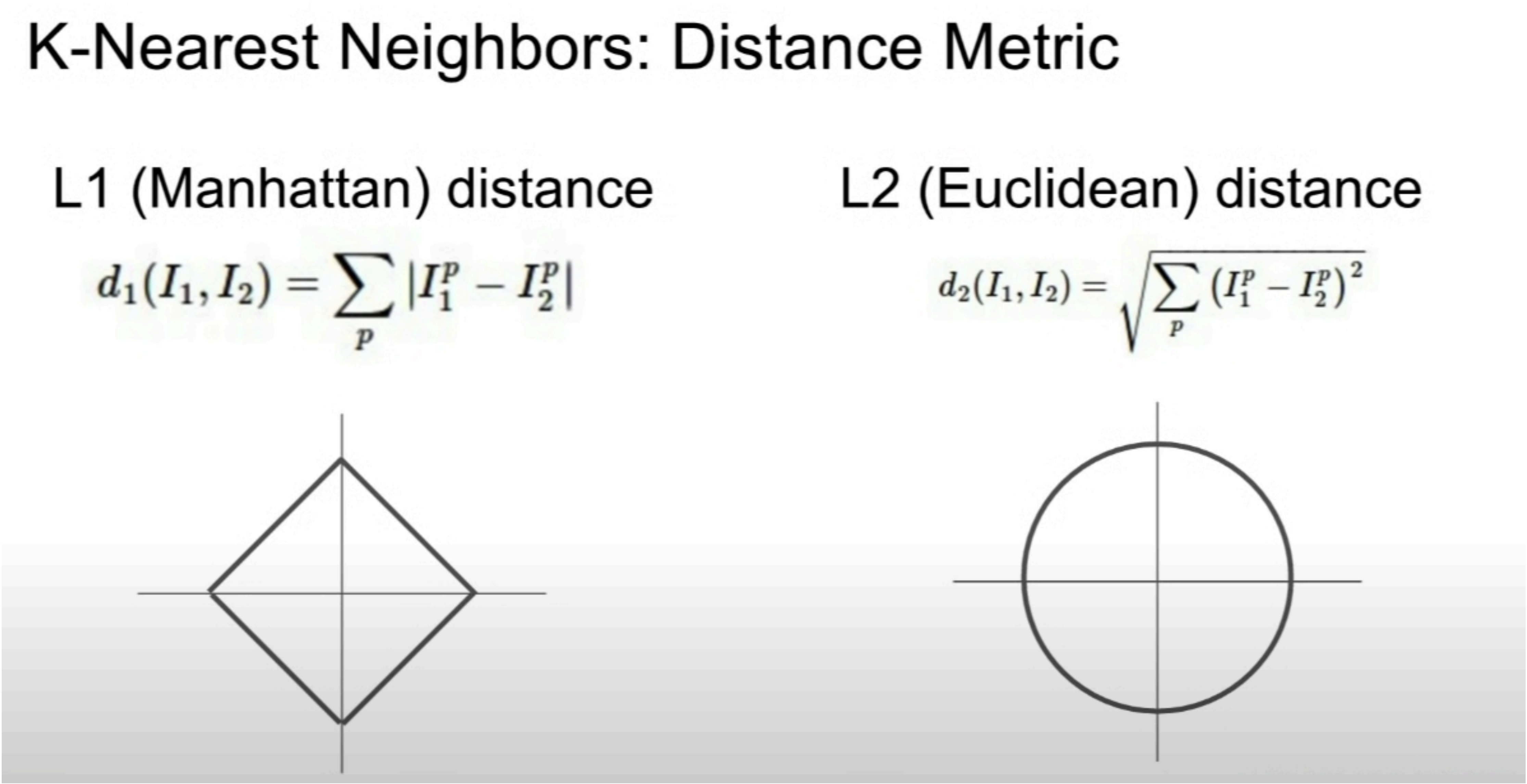
Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N} (pred_i - target_i)^2}$$

Introduction

- Cs231n 2강에서 다룬 내용이기는 하지만, 이때 강의자의 L1 distance, L2 distance 를 언제 사용하는지에 대한 설명이 분명하게 이해되지 않았었다.

L1 vs. L2. It is interesting to consider differences between the two metrics. In particular, the L2 distance is much more unforgiving than the L1 distance when it comes to differences between two vectors. That is, the L2 distance prefers many medium disagreements to one big one



Introduction

- 결과 미리보기

	MAE	MSE	RMSE
이상치에 대한 민감도	비교적 낮음	매우 높음	비교적 높음
미분 가능 여부	원점에서 불가	모든 점에서 가능	**원점에서 불가**

Main

Epoch	Prediction	Target
	모델이 예측한 값	정답 값
1	[0, 4, 9]	[3, 5, 7]
2	[2, 4, 2]	[3, 5, 7]
3	[3, 5, 6]	[3, 5, 7]

Outlier

$$MSE = \frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2}$$

$$MAE = \frac{1}{N}\sum_{i}^{N}|(pred_i - target_i)|$$

Main

$MSE = \frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2$	$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2}$	$MAE = \frac{1}{N}\sum_{i}^{N}\lvert(pred_i - target_i)\rvert$
$E1 : \frac{1}{3}(9+1+4) = \frac{14}{3}$	$E1 : \frac{\sqrt{42}}{3}$	$E1 : 2$
$\boxed{E2 : \frac{1}{3}(27) = 9}$	$E2 : 3$	$\boxed{E2 : \frac{7}{3}}$
$E3 : \frac{1}{3}$	$E3 : \frac{\sqrt{(3)}}{3}$	$E3 : \frac{1}{3}$

이상치에 민감

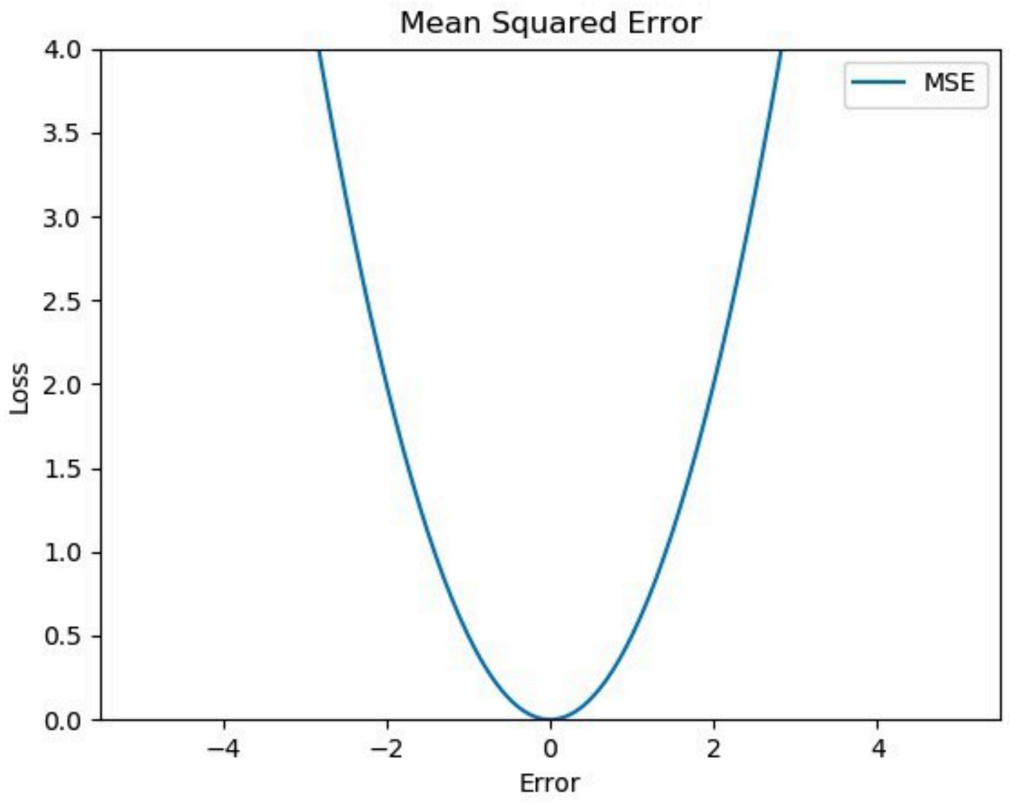
이상치에 둔감

MSE 특징

[1] Mean Square Error는 예측값과 정답의 차이를 제곱하기 때문에, 이상치에 대해 민감하다. 즉, 정답에 대해 예측값이 매우 다른 경우, 그 차이는 오차값에 상대적으로 크게 반영된다.

[2] 오차값에 제곱을 취하기 때문에 (1) 오차가 0과 1 사이인 경우에, MSE에서 그 오차는 본래보다 더 작게 반영되고, (2) 오차가 1보다 클 때는 본래보다 더 크게 반영된다.

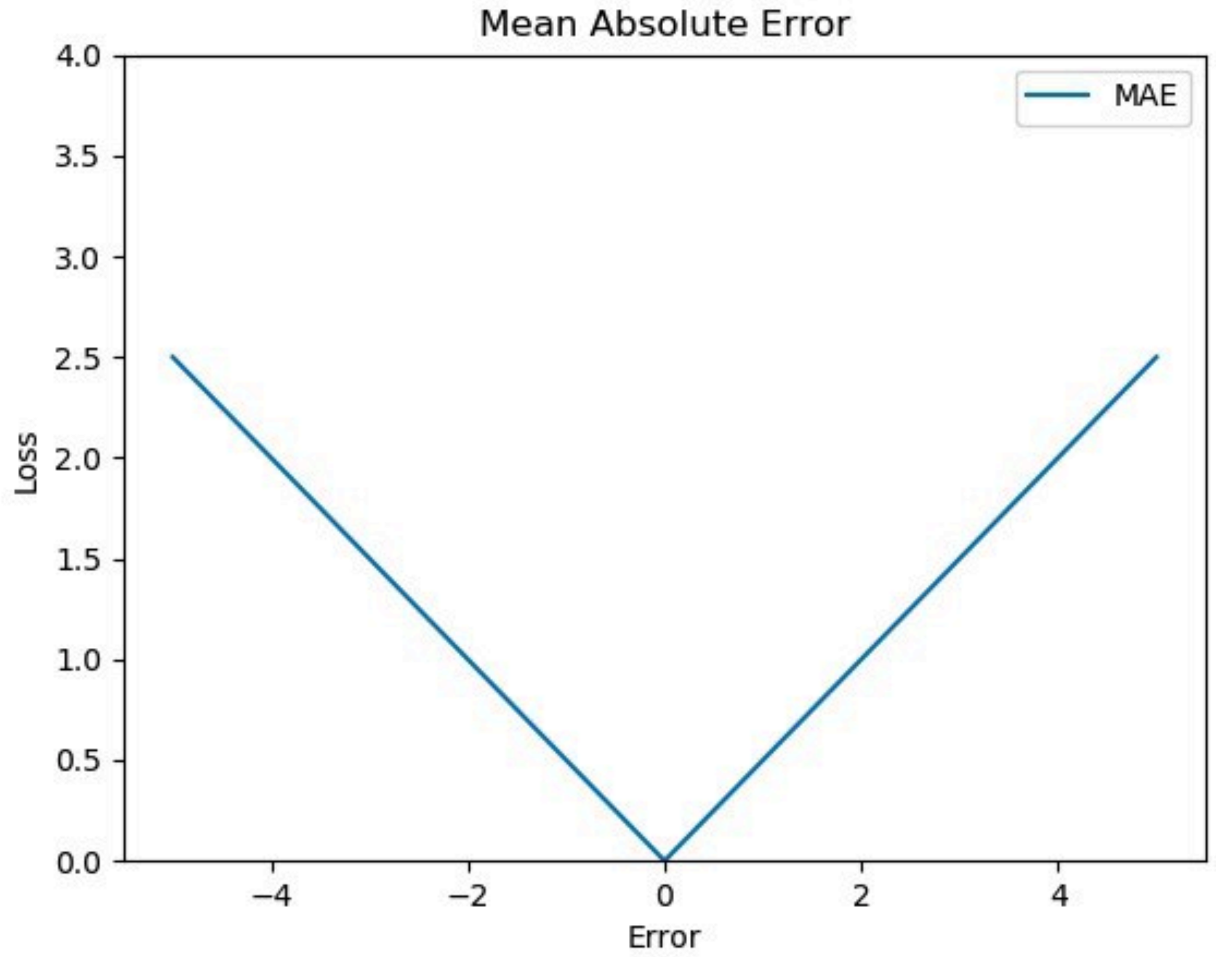
[3] 모든 함수값이 미분 가능하다. MSE는 이차 함수이기 때문에 아래와 같이 첨점을 갖지 않는다.



MAE 특징

[1] 이상치에 둔감 혹은 강건robust하다. 그 이유는 Figure 1에서 보듯이, (3)에 해당하는 MAE는 위 MSE, RMSE에 비해, 오차값이 outlier의 영향을 상대적으로 크게 받지 않는다.

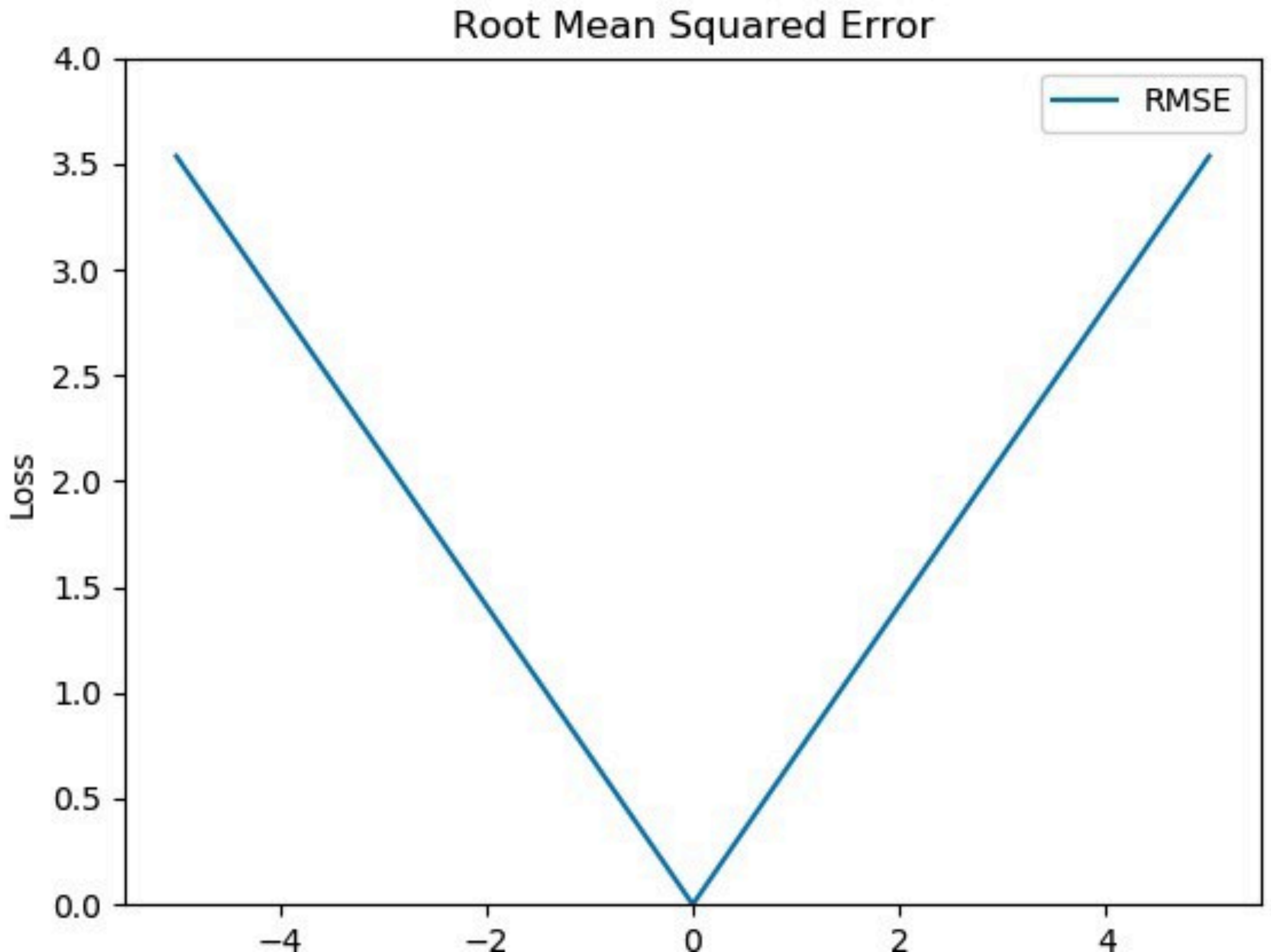
[2] 함수값에 미분 불가능한 지점이 있다. 아래 Figure 3의 MAE를 표현한 함수의 최솟값은 첨점이기 때문에 미분이 불가능하다.



RMSE 특징

[1] MSE에서 루트를 취하기 때문에, MSE 특징의 [1], [2] 에서 나타나는 단점이 어느 정도 해소된다.

[2] MSE는 부드러운 곡선형으로 오차 함수가 그려지지만, RMSE는 그 MSE에서 루트를 취하기 때문에 미분 불가능한 지점을 갖는다.



RMSE vs MAE

- **MSE에 루트를 취하여 이상치에 대한 민감도를 줄인 RMSE와 절대값을 취하는 MAE 는 어떤 차이가 있는가?**

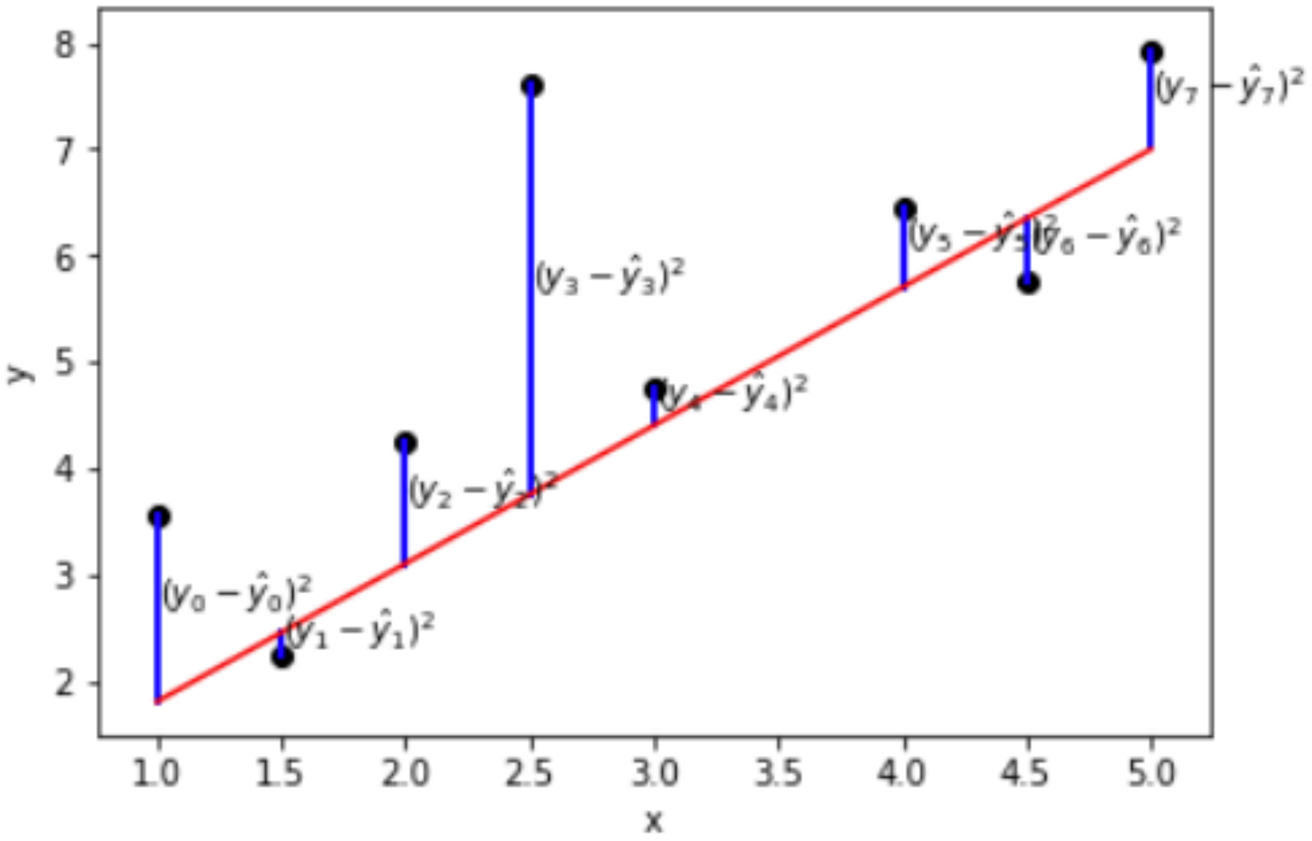
→ MAE는 오차들의 절댓값의 평균을 계산한다는 점에서, 모든 examples에 대한 오차에 동일한 가중치를 부여

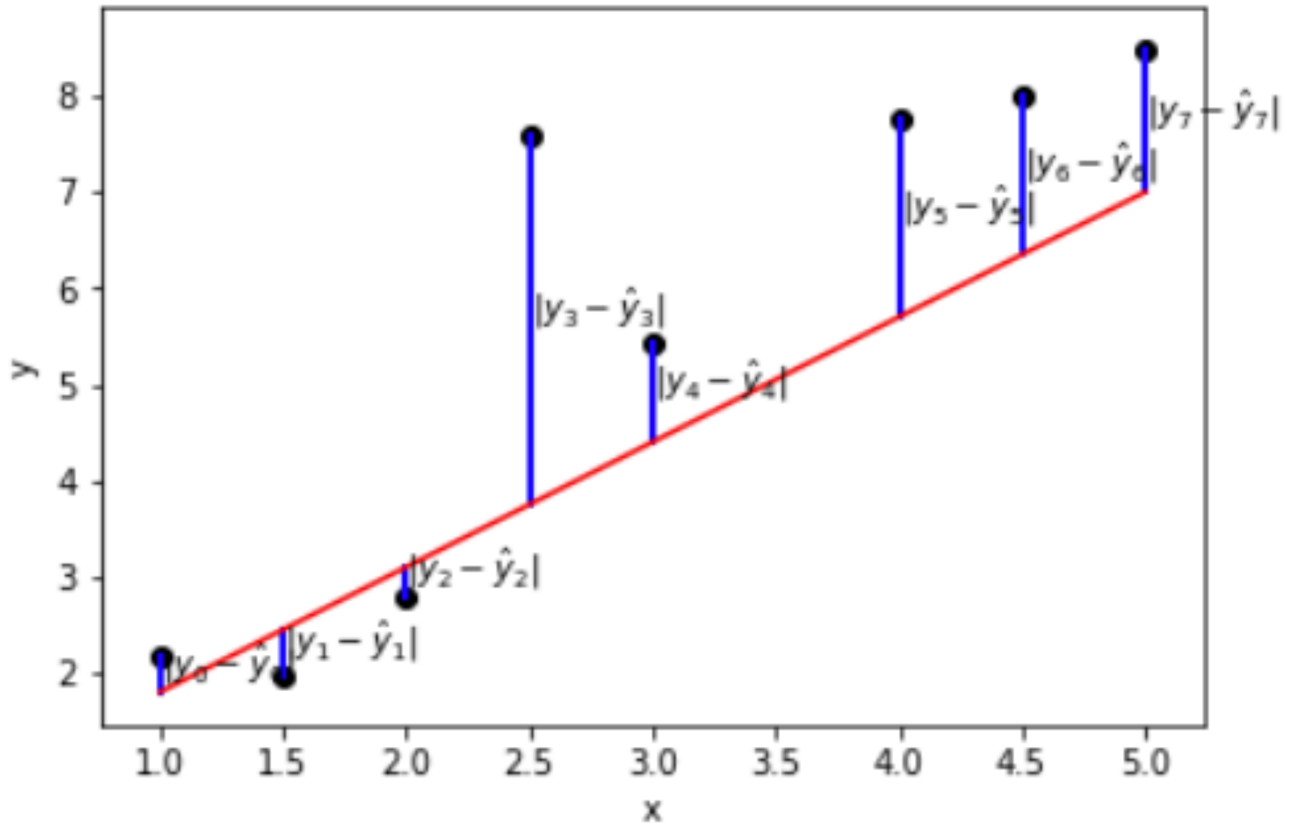
→ RMSE는 각 error에 제곱을 취한 뒤 평균을 구하고, 그것에 루트를 씌우는 것이기 때문에, 각 오차가 다른 가중치를 가짐

제곱을 취하는 과정에서, 에러가 큰 경우는 더 가중(ex. err=4 -> 16)되고, 에러가 작은 경우(ex. err=1.1 -> 1.21)는 덜 가중된다.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2}$$

$$MAE = \frac{1}{N}\sum_{i}^{N}|(pred_i - target_i)|$$





The RMSE is differentiable?

- 일부 블로거[7],[8] 등은 RMSE는 미분 가능한 함수라고 말한다.

> *If we define loss function (J) in terms of RMSE: then we can easily differentiate J wrt. to m and b and get the updated m and b*

- 하지만 내 생각에 RMSE는 제곱 항의 합에 루트를 취하기 때문에 원점에서 미분 불가능한 것 같다.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2} \longrightarrow F(pred_i, target_i) = \sqrt{\frac{1}{2}\sum_{i}^{2}(pred_i - target_i)^2} \longrightarrow F(x, y) = \frac{1}{\sqrt{2}}\sqrt{x^2 + y^2}$$

where $x = (pred_1 - target_1)$

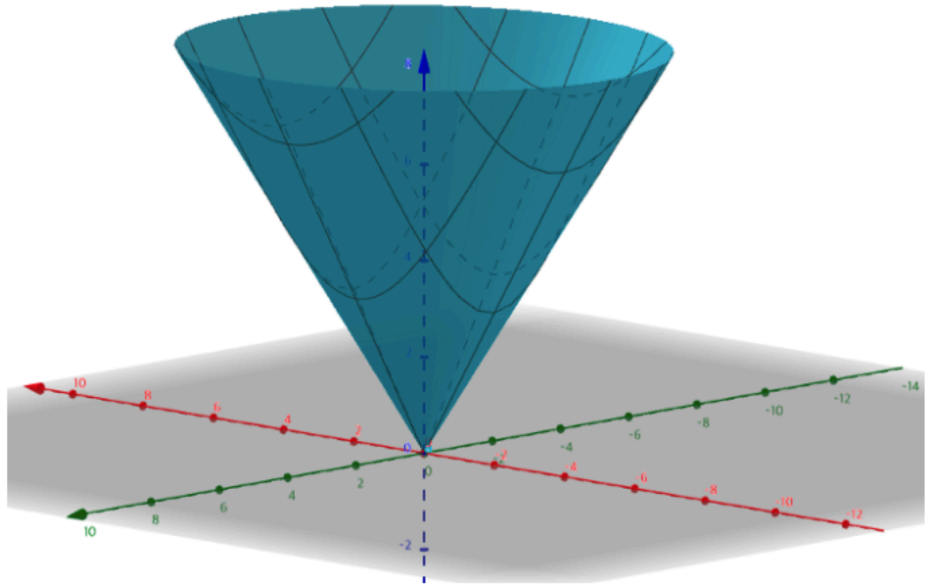
The RMSE is differentiable?

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(pred_i - target_i)^2} \longrightarrow F(pred_i, target_i) = \sqrt{\frac{1}{2}\sum_{i}^{2}(pred_i - target_i)^2} \longrightarrow F(x, y) = \frac{1}{\sqrt{2}}\sqrt{x^2 + y^2}$$

where $x = (pred_1 - target_1)$

Definition Differentiable: Two Variables Let $f: \mathbb{R}^2 \to \mathbb{R}$. We say f is ***differentiable*** at (x_0, y_0), if $\partial f/\partial x$ and $\partial f/\partial y$ exist at (x_0, y_0) and if

$$\frac{f(x, y) - f(x_0, y_0) - \left[\frac{\partial f}{\partial x}(x_0, y_0)\right](x - x_0) - \left[\frac{\partial f}{\partial y}(x_0, y_0)\right](y - y_0)}{\|(x, y) - (x_0, y_0)\|} \to 0 \tag{2}$$



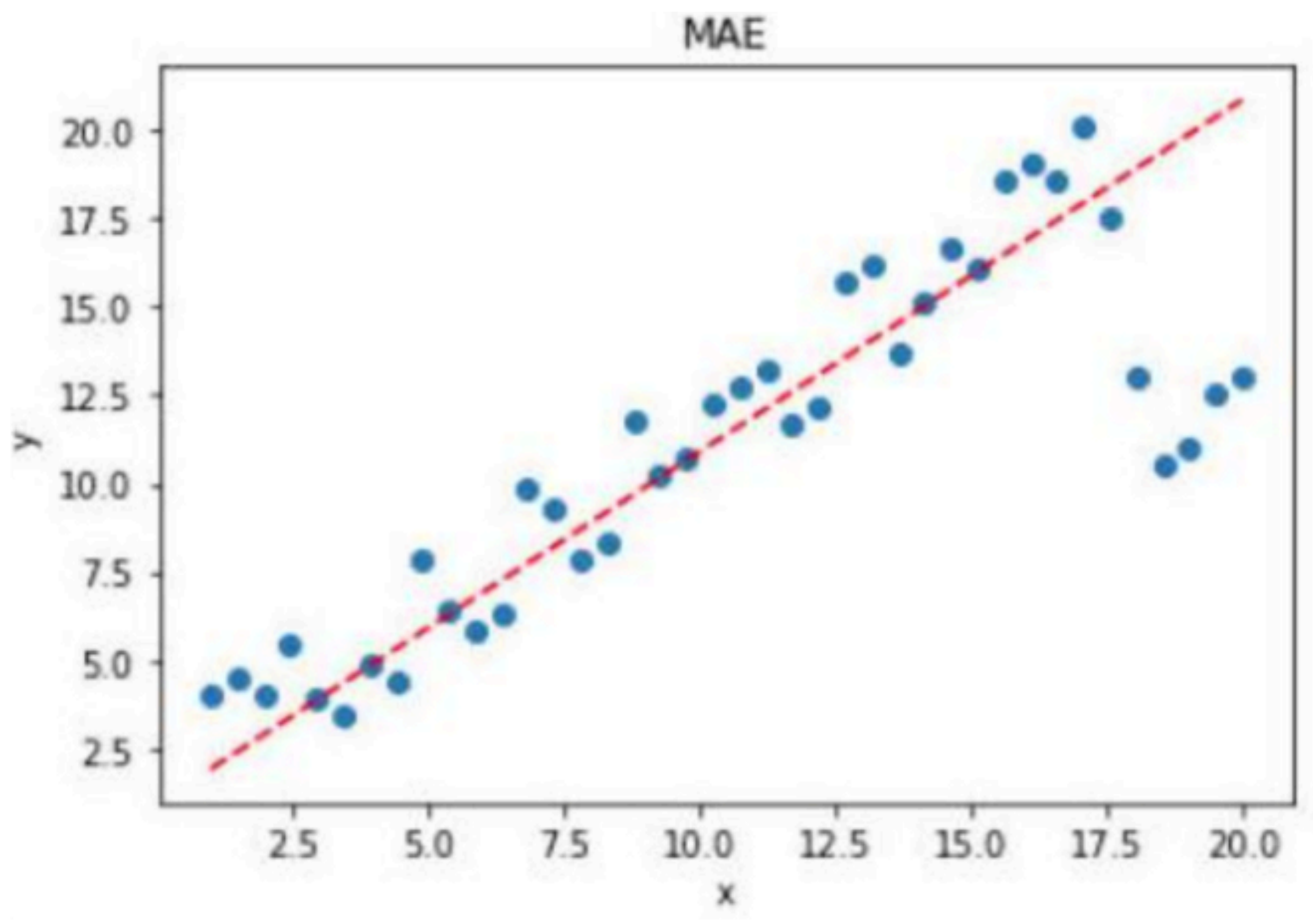
언제 어떤 loss를 사용하는가

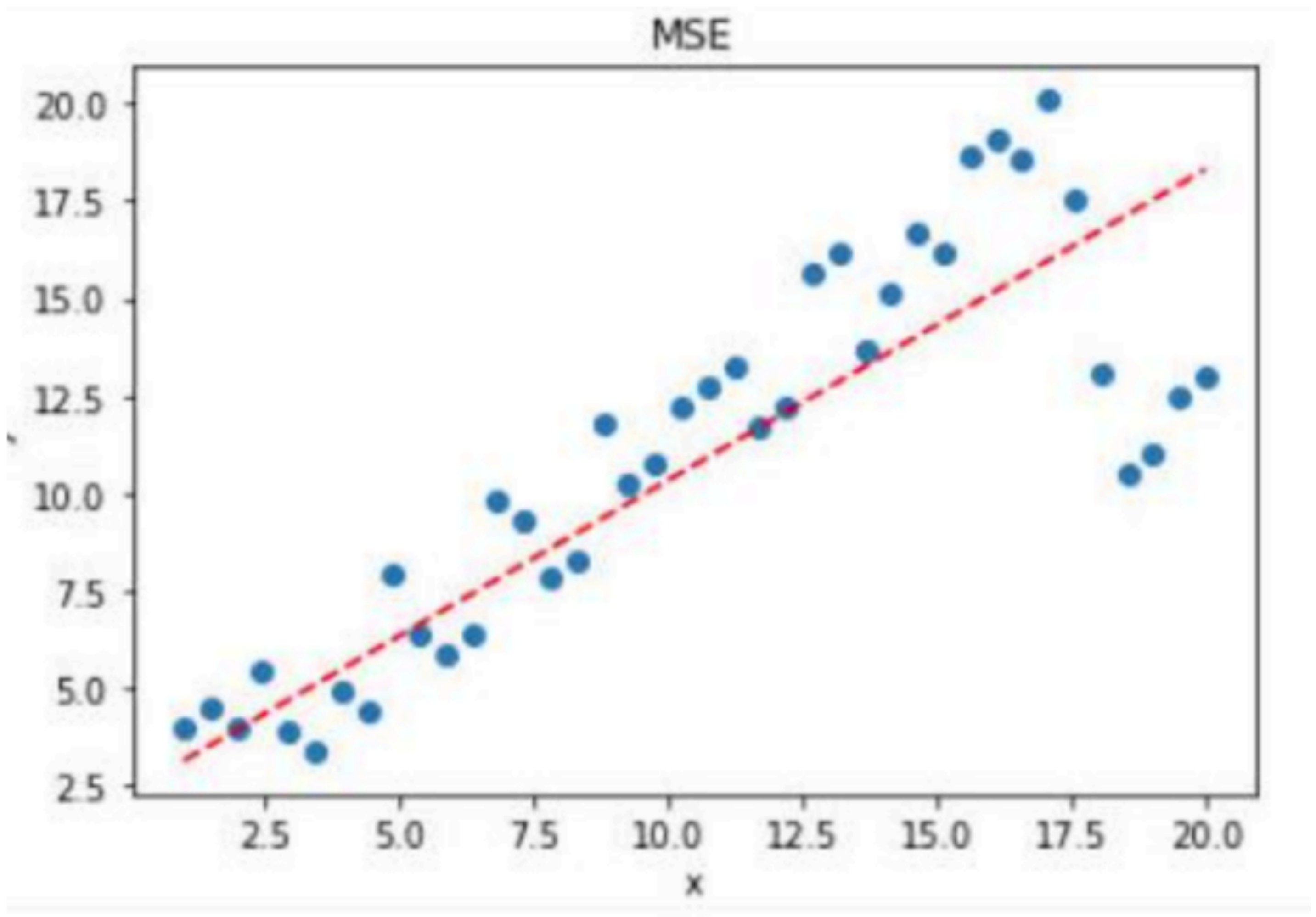
	MAE	MSE	RMSE
이상치에 대한 민감도	비교적 낮음	매우 높음	비교적 높음
미분 가능 여부	원점에서 불가	모든 점에서 가능	원점에서 불가

- 이상치의 영향을 덜 받으며 모델을 만들고자 할 때 사용
- 소수의 특정 이상치에 적응하며 학습을 하지 않고 일반적인 경향을 학습

- 이상치에 비교적 민감하기 때문에, 이상치를 중하게 고려하여 모델링 할 필요가 있을 때 사용
- 소수의 특정 이상치에도 적응하며 학습하기 때문에 MAE 보다 (이상치도 중하게 고려한다는 의미에서) 더 전체적인 경향성을 파악하며 학습

언제 어떤 loss를 사용하는가





References

https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e

https://dailyheumsi.tistory.com/167

https://medium.com/analytics-vidhya/a-comprehensive-guide-to-loss-functions-part-1-regression-ff8b847675d6

https://m.blog.naver.com/alwaysneoi/100135882596

https://jmlb.github.io/flashcards/2018/07/01/mae_vs_rmse/

ttps://devstarsj.github.io/data/2018/10/22/kaggle.coursera.competition.03.01/

https://discuss.analyticsvidhya.com/t/statistics-why-squaring-is-used-to-calculate-error-in-rmse/14163

https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4

https://deep-learning-study.tistory.com/663