Loss functions

27 September 2023 07:06

Data → NN → $\hat{y}$ → Loss ← [WB] → optimizer

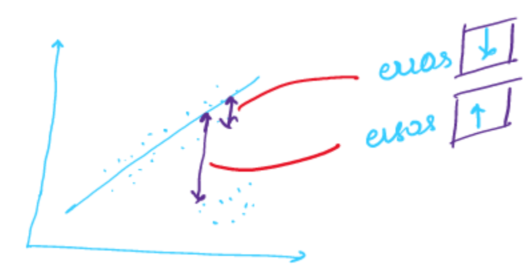
$y_a - \hat{y}$

forward propagation ←

Backward propagation

You can't improve what you can't measure

Loss function is an method of evaluating how well our algorithm is performing on given Data set.



→ Regression
→ classification

Loss $y_{act} - \hat{y}$

Loss [↑] perf [↓]

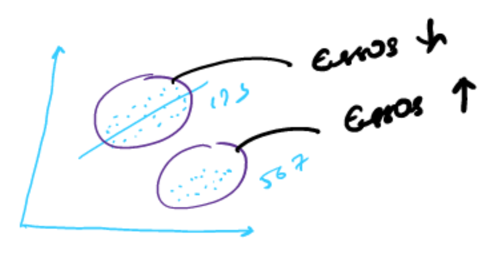
Loss [↓] perf [↑]

Loss function = single record
cost function - whole batch

$y_{act} - \hat{y}$

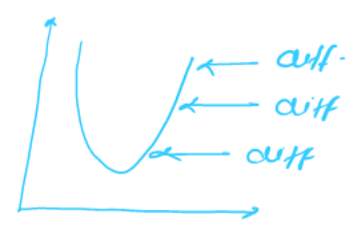
Mean squared Error.

$$\frac{1}{n}\sum_{i=1}^{n}(y_{act} - \hat{y})^2$$



Error.
1 — 1
2 — 4
3 — 9
5 — 25
6 — 36
7 — 49
15 — 255

MSE →

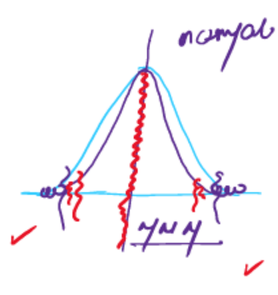


Advantages.

1. it is in quadratic form where we can plot gradient descent with only one global minima.
2. Easy to interpret.
3. differentiable.

Disadvantages.

1. Error unit is not same.
2. not Robust to outlier.



true nature of Error is not being captured.

Mean Absolute Error.

Loss = $|y_{act} - \hat{y}|$ = Error ← linear

Cost = $\frac{1}{n}\sum_{i=1}^{n}|y_{act} - \hat{y}|$ ← linear

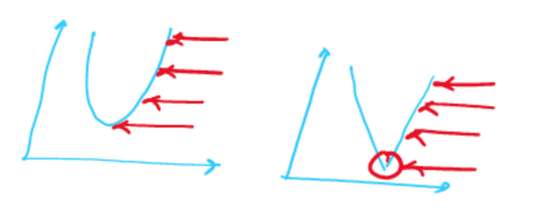
Error.
1 — 1
2 — 2
3 — 3
5 — 5
6 — 6
7 — 7
15 — 15

Advantages.

1. Easy to interpret
2. same unit as output.
3. outliers are better handled here than MSE as it is not penalizing the model by squaring the Errors.

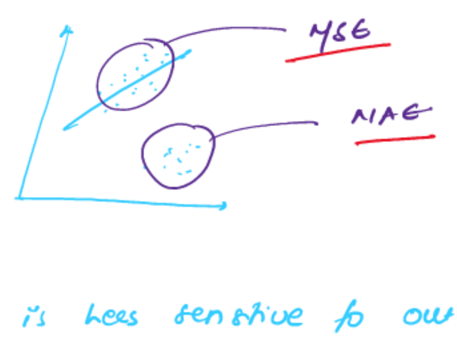
Disadvantages.

1. graph is not differentiable
2. multiple local minima.



Huber Loss.

$$\text{Loss} = \begin{cases} \frac{1}{2}(y_{act} - \hat{y})^2 & \leftarrow |y_{act} - \hat{y}| \leq \delta \\ \delta|y_{act} - \hat{y}| - \frac{1}{2}\delta^2 & \leftarrow > \delta \end{cases}$$



Errors
MSE: 1 — 1, 2 — 4, 3 — 9
MAE: 5 — 5, 6 — 6, 7 — 7.

it is less sensitive to outliers.

whenever the data point is an outlier huber loss will behave like MAE

whenever the data point is near to Regression line huber loss will behave like MSE

MSE → MAE, quad → linear } δ.

MAE → MSE, linear → quad. } δ

Advantage.

1. outliers are better handled here
2. Local minima situations are also better handled

Disadvantages

1. complex.

Mean squared logarithmic Error.

Data → NN → $\hat{y}$ → Loss ← WB.

$\hat{y}$ → $\log(\hat{y})$ → $(y_{act} - \hat{y}_{log})^2$

natural log.