WHAT YOU WILL STUDY IN TODAY VIDEO ?

Prerequistes. {MLP.}

- What is a Loss Function?
- Types of Loss Functions
- Formula , Derivative , Use Case of Each Loss Function
- Problems of Each Loss Functions

Introduction

* Difference actual value (y) & predicted value ($\hat{y}$)

$$\boxed{y - \hat{y}}$$

* Model performance

$5 \to 4$ (1) $5 \to 3$ (2)

Lower loss = Greater performance

Imp.

Loss $\neq$ Accuracy

$\begin{Bmatrix} 0.97 \\ 0.67 \\ 7.63 \end{Bmatrix}$ Continous value

Percentages (90%)

Total no of correct prediction

how far Predicted values are from actual values

* Model → low loss ✓
→ low accuracy

* Loss main aim → To guide Model Training. (By updating weights)

Types.

Regression

1. MSE
2. MAE
3. Huber Loss

Classification

4. Binary Cross entropy
5. Imp Categorical Cross entropy
6. Sparse Categorical Cross entropy

① MSE.

↳ Mean Squared error

formula.

$$\boxed{L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

eg. Actual values $= [3, 4, 5]$ Pred values $= [2.5, 3.8, 5.5]$

$$L = \frac{1}{3}\left[(3-2.5)^2 + (4-3.8)^2 + (5-5.5)^2\right]$$

$$L = 0.2067$$

Derivative

$$\frac{dL}{d\hat{y}} = \frac{-2}{n}\sum(y-\hat{y})$$

Use ⟶ Regression :- O/p feature is continous
eg. House Prediction Price

Problem. ⟶ Sensitive to outliers
(Make error too big)

Problem $\longrightarrow$ Sensitive.
(Make error too big)

② MAE. → Mean absolute error

formula

$$\boxed{L = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|}$$

eg. Actual = [3, 4, 5]   Pred = [2.5, 3.8, 5.5]

$$L = \frac{1}{3}\left[(|3-2.5|) + (|4-3.8|) + (|5-5.5|)\right]$$

$$L = 0.4$$

Derivative

$$\boxed{\frac{dL}{d\hat{y}} = \frac{1}{n}\sum \text{sign}(y-\hat{y})}$$

sign → signum fn.

$$\therefore \text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

eg.

Actual	Pred.	$y-\hat{y}$	$\text{sign}(y-\hat{y})$
5	3	2	1
7	8	-1	-1
6	6	0	0

$$\frac{dL}{d\hat{y}} = \frac{1}{3}\left[1 + (-1) + 0\right]$$

$$= \underline{\underline{0}}.$$

Use $\longrightarrow$ Regression $\longrightarrow$ Robust to outliers

Problem. $\longrightarrow$ Gradient is not smooth at $y = \hat{y}$

means

loss fn is not differentiable at $y = \hat{y}$

$\downarrow$

optimization takes time

↓

Optimization takes time

Slower convergence

→ (MSE don't have this problem)

③ Huber Loss

→ Combination of MSE & MAE.

parameter delta (δ)

MSE → small error
MAE → large error

formula

$$L = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & , \text{ if } |y-\hat{y}| \leq \delta \\ \delta|y-\hat{y}| - \frac{1}{2}\delta & , \text{ if } |y-\hat{y}| > \delta \end{cases}$$

eg. $\delta = 1.$ Actual = 3 Pred = 1

$|y-\hat{y}| = |3-1| = 2.$

$\therefore 2 > \delta$

$L = 1 \times 2 - \frac{1}{2} \times 1$

$= 2 - 0.5$

$= 1.5.$

$= 1.5$

Derivative.

$$\frac{dL}{d\hat{y}} = \begin{cases} \overset{MSE}{(y - \hat{y})} & , \text{if } |y-\hat{y}| \leq \delta \\ \delta \cdot \overset{MAE}{\text{sign}(y - \hat{y})} & , \text{if } |y-\hat{y}| > \delta \end{cases}$$

Use → Regression
Robust to Outlier ✓
Differentiable at $y = \hat{y}$ ✓

Problem → δ
Tuning of finding best δ value.

4) Binary Cross Entropy

formula.

$$L = \frac{-1}{n} \sum \left[y \log(\hat{y}) + (1-y) \log(1-\hat{y}) \right]$$

Derivative.

$$\boxed{\frac{dL}{d\hat{y}} = \frac{\hat{y}-y}{\hat{y}(1-\hat{y})}} =$$

Use. ⟶ Binary Classification

Problem ⟶ Only binary classfication ×

(5) Categorical Cross Entropy

formula

$$\boxed{L = -\sum y \log(\hat{y})}$$

3 classes

eg:

Actual = [1 , 0 , 0]

Pred = [0.8 , 0.1 , 0.1]

$$L = -\left[1 \log(0.8) + 0 \log(0.1) + 0 \log(0.1)\right]$$

$$= -\log(0.8)$$

$$= 0.223 \longrightarrow \text{Loss}$$

Derivative

$$\boxed{\frac{dL}{d\hat{y}} = \frac{-y}{\hat{y}}}$$

Use $\longrightarrow$ Multi class classification

Use → Multiclass ca

Problem → Not memory efficient

Why: due to vectors (one hot encoded)

Actual = setosa

setosa	versicolor	virginica

[1 , 0 , 0] → memory efficient ✗

6 Sparse Categorical Cross Entropy

Two scenario

$\boxed{\text{Actual} = \text{setosa}}$

Vectors

[1, 0, 0]

one hot encoding

Categorical Cross entropy

0 → setosa
1 → versicolor
2 → virginica

Indices

0

label encoding

memory efficient

Sparse categorical Cross entropy