

WHAT YOU WILL STUDY IN TODAY VIDEO ?

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

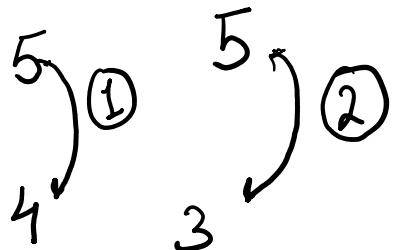
Prerequisites
MLP.

Introduction

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

$$y - \hat{y}$$

- + Model performance



Lower loss = Greater performance

Imp.

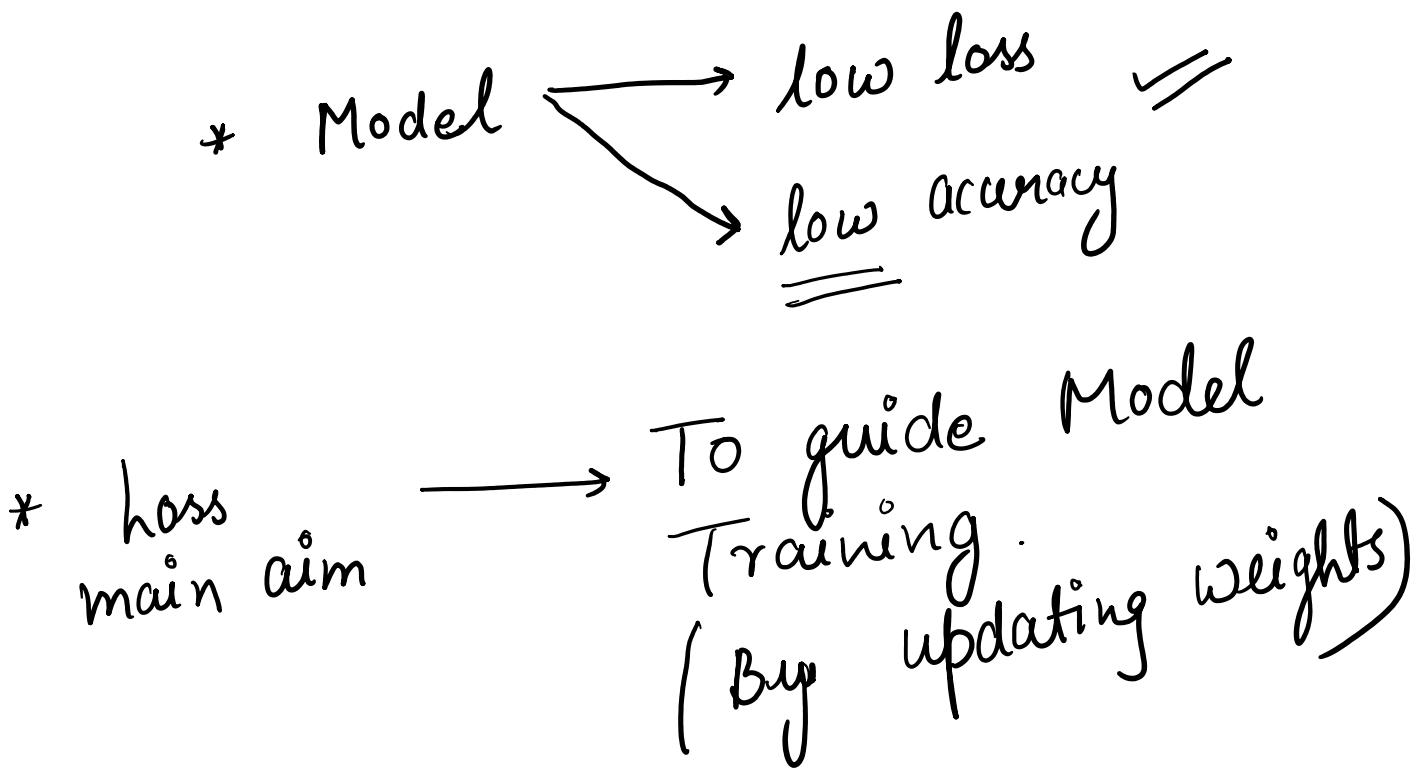
Loss \neq Accuracy

Percentages (90%)

{
0.97
0.67
1.63}
Continuous value

how far Predicted values
are from actual values

Total no of
correct prediction



Regression

① $MSE =$

② $MAE =$

③ Huber loss

~~Types~~

Classification

④ Binary Cross entropy =

⑤ Categorical Cross entropy =

⑥ Sparse Categorical Cross entropy =

① MSE

↳ Mean Squared error

formula:

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

eg:
~~if~~ Actual values = $[3, 4, 5]$

Pred values = $[2.5, 3.8, 5.5]$

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

$$L = 0.2067 =$$

Derivative

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

Use → Regression :- O/P feature is continuous
eg. House Prediction
Price

problem → Sensitive to outliers
(Mse. error too big)

problem → Sensitive
(Make error too big)

② MAE

formula:

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

eg.

$$\text{Actual} = [3, 4, 5]$$

$$\text{Pred} = [2.5, 3.8, 5.5]$$

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

$$\underline{\underline{L = 0.4}}$$

Derivative

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

sign \rightarrow Signum
fns.

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

$\wedge \quad \dots (u - \hat{u})$

<u>eg.</u>	<u>Actual</u>	<u>Pred.</u>	$y - \hat{y}$	<u>sign</u> ($y - \hat{y}$)
	5	3	2	1
	7	8	-1	-1
	6	6	0	0

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

use → Regression → Robust to Outliers

problem: → Gradient is not smooth at $y = \hat{y}$

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

↓
optimization takes time
... n.

Optimization takes time
↓
Slower convergence

→ (MSE don't
have this problem)

③ Huber Loss

parameter
delta
(δ)

Combination of MSE & MAE
 {
 MSE → small error
 MAE → large error
 }

formula

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

e.g.

$$\delta = 1 \quad \text{Actual} = 3$$

$$P_{\text{pred}} = 1$$

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

$$\therefore 2 \geq \underline{\delta}$$

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

$$= 2 - 0.5$$

$$= \underline{\underline{1.5}}$$

Derivative:

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

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 classification \times

⑤ Categorical Cross Entropy

formula

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

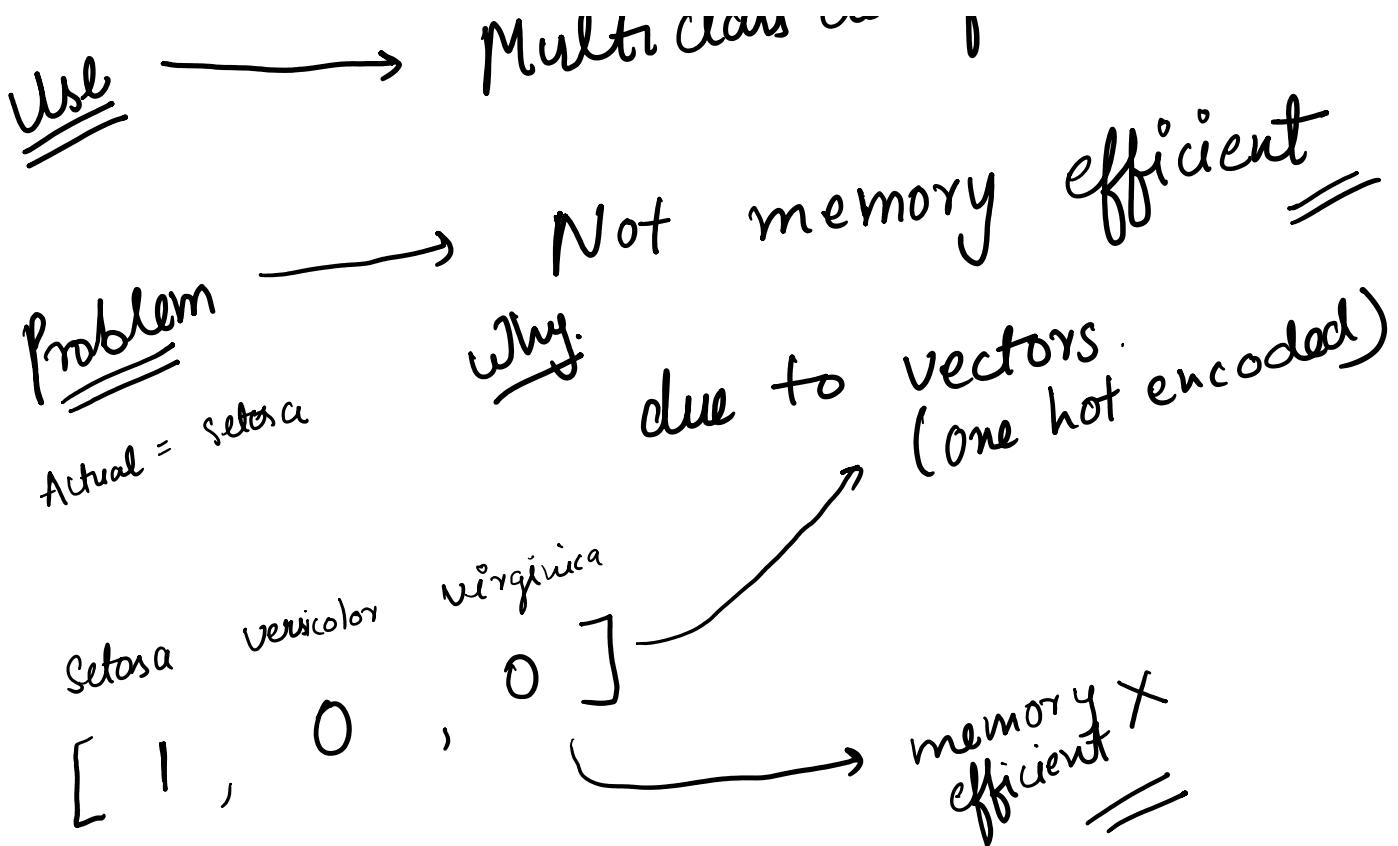
3 classes
eg. Actual = $\begin{bmatrix} 1, 0, 0 \end{bmatrix}$
 Pred = $\begin{bmatrix} 0.8, 0.1, 0.1 \end{bmatrix}$

$$\begin{aligned} L &= - \left[1 \log(0.8) + 0 \log(0.1) + 0 \log(0.1) \right] \\ &= - \log(0.8) \\ &= \underline{0.223} \rightarrow \underline{\text{loss}} \end{aligned}$$

Derivative

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

use \longrightarrow Multi class classification



⑥ ~~Sparse~~ Categorical Cross Entropy

