→ What is Loss function?

Loss function is a method of ~~en~~ evaluating how well your algorithm is modelling your dataset.

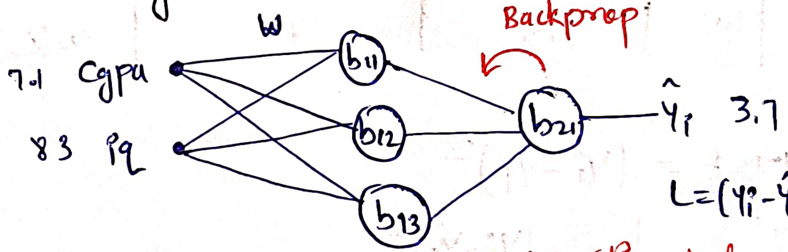
if loss function → high (algorithm is poor)
if loss function → low (algorithm is great)

Why is Loss function important?

You can't improve what you can't measure

Loss function in Deeplearning.

cgpa	iq	package
7.1	83	3.2
8.5	91	4.5
6.3	102	6.1
5.1	87	2.7



$L = (y_i - \hat{y}_i)^2 = (3.2 - 3.7)^2$

According to loss function it is going back and arjest the value of weights and bieah.

Loss functions in DL

Regression
- MSE
- MAE
- nuber loss

Classification
- binary crossentropy
- categorical cross entro
- hinge loss

Autoencoders
- KL divergence

GAN
- discriminator loss
- Minmax gan loss

Object detection
- Focal loss

Embedding
- Triplet loss

→ **Loss function Vs Cost function:**

Loss function OR Error function → Single training ka loss.

$(y_i - \hat{y}_i)^2$. Eg :- $y_i = 6.3$, $\hat{y}_i = 6.1$

Lossfun^n = $(6.3 - 6.1)^2$

Cost function:- All over data ka loss is called cost function.

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

→ **Some loss functions**

Regression

1. Mean Squared Error (MSE) :-
   OR Squared loss OR L2 Loss

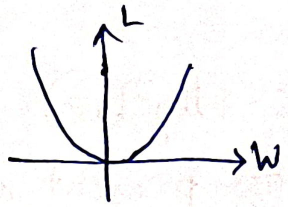
cgpa	iq	package (y_i)	predication ($\hat{y}_i$)	$y_i - \hat{y}_i$
6.3	100	6.3	6.1	0.2
7.1	91	4.1	4	0.1
8.5	83	3.5	3.7	−0.2
9.2	102	7.5	7	0.2

Jiska $(y_i - \hat{y}_i)$ high hoga uske karen weight ~~or~~ & biases mai drastic change aya gai.
Jisk ka low hai usmai kaam change hoga.

$$L_{\cancel{MSE}} = (y_i - \hat{y}_i)^2$$

$$C = \frac{1}{n}\sum (y_i - \hat{y}_i)^2$$

Advantagys



1. Easy to Intepret.
2. Diffrentiable
3. Only 1 local Minima.

Disadvantagys

1. Error unit (squared)
2. Not Robut to outlier.

\# Outlier ko jyada value krta hai

\# If their is too many outlier then MSE is not used.

\# Output ~~last~~ layear mai activation function linear hona chiya then we use → MSE.

2. Mean Absolute Error (MAE)
   OR L1 loss.

$L = |y_i - \hat{y}_i|$

$C = \frac{1}{n} \Sigma |y_i - \hat{y}_i|$

\# If their is outlier in data then we use MAE.

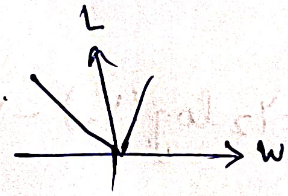
Advantages

1. Intuitive and easy to understand.
2. Unit → same → y
3. Robust to outliers

\# Outlier ko jyada value nahi krta hai.

Disadvantagys

1. Not diffrentiable (Subgradient is used)



3. Huber Loss

$$L = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & \text{for } |y-\hat{y}| \leq \delta \rightarrow \text{outlier nahi hai} \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise} \rightarrow \text{outlier hai} \end{cases}$$

δ → hyperparamiter

\# Combination of both MSE and MAE.

Classification

1. Binary Cross Entropy :- (log loss)

When Two class present only → like placed or not.

Loss funⁿ = $-y \log(\hat{y}) - (1-y) \log(1-\hat{y})$

$y \rightarrow$ actual value

$\hat{y} \rightarrow$ prediction

In output layer the activation function is Sigmoid then we use BCE.

Cost funⁿ = $-\frac{1}{n}\left[\sum_{i=0}^{n} y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)\right]$

Advantage

1. Differentiable

Disadvantage

1. Multi local minuma
2. intuitive.

2. Categorical Cross Entropy :-

↳ when multi-class Classification problem.

↳ Used in softmax Regression.

$$L = -\sum_{j=1}^{k} y_j \log(\hat{y}_j)$$ where k is no. of classes.

let $k=3$

$$L = -y_1 \log(\hat{y}_1) - y_2 \log(\hat{y}_2) - y_3 \log(\hat{y}_3)$$

Used for slow categories.

Activation funⁿ of output layer → soft max.

$$f(z_1) = \frac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}}, \quad f(z_2) = \frac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}}, \quad f(z_3) =$$

$$f(z_1) + f(z_2) + f(z_3) = 1$$

In output layer no. of node = no. of classes.

output column per → one hot encoding.

$$C = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij} \log(\hat{y}_{ij})$$

Sparse categorical Cross Entropy

In sparse categorical output column is ~~used~~ converted → Integer encoding.

All is same like ~~spar~~ categorical Cross entropy

when $\underline{Y=1}$ predication → $[0.1 \quad 0.4 \quad 0.5]$ $Y=1$ ← 1/2/3

$$L = -Y_1 \log(\hat{Y}_1) = -1 \log(0.1)$$

fast then categorical cross

when $\underline{Y=2}$

$$L = -Y_2 (\log \hat{Y}_2) = -1 \log(0.4)$$

Used in large caterogical.