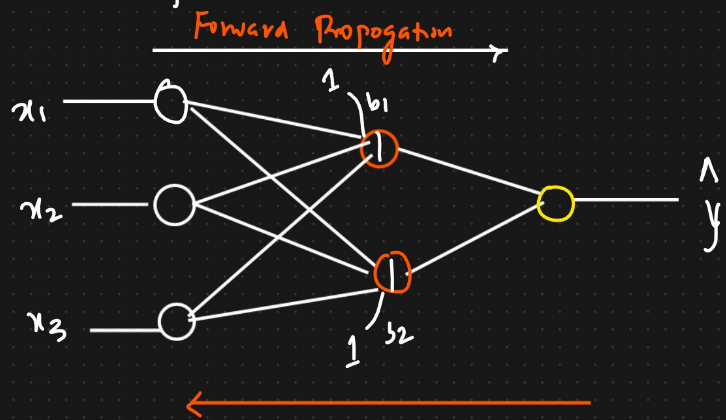
Loss function And Cost Function



$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}\overbrace{(y_i-\hat{y}_i)^2}^{\text{Error}} \downarrow\downarrow\downarrow$$

$$\boxed{\text{Loss} = (y-\hat{y})^2} \downarrow\downarrow\downarrow$$

Optimizers

Gradient Descent

x_1	x_2	x_3	$\hat{y}$ O/P
–	–	–	0
–	–	–	1
–	–	–	0
–	–	–	1

Loss function

$$MSE = (y-\hat{y})^2$$

Cost function

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

ANN

Classification

Regression

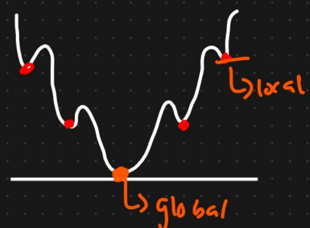
① Mean Squared Error (MSE)
② Mean Absolute Error (MAE)
③ Huber Loss
④ RMSE

① Mean Squared Error (MSE)

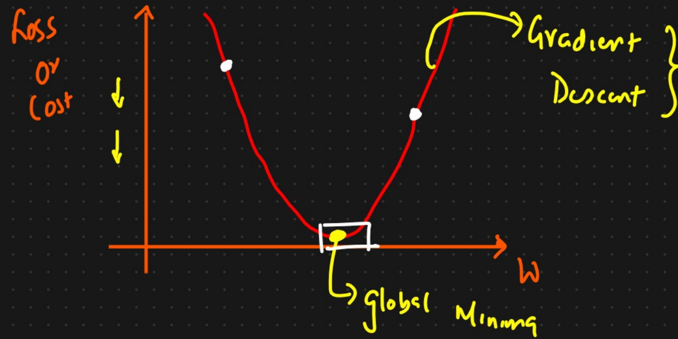
$$\text{Loss function} = (y-\hat{y})^2$$

↳ Quadratic Equations

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



Weight updation



$$W_{new} = W_{old} - \eta \left[\boxed{\frac{\partial L}{\partial W_{old}}} \right]$$

$\Uparrow$ Slope

Advantages

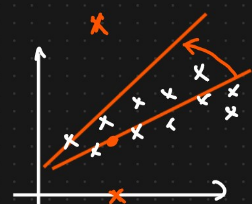
① MSE is Differentiable

② It has 1 local or global Minima

③ It converges faster

Disadvantages

① Not Robust to outliers



penalizing the Error

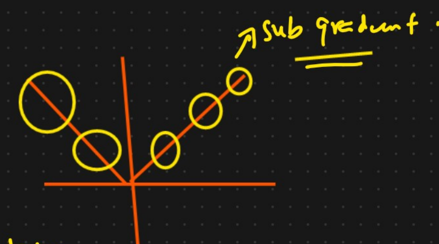
↗ Cost function =

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

② Mean Absolute Error (MAE)

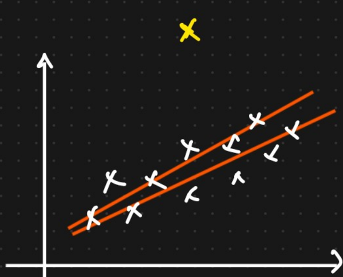
Loss fn : $|y - \hat{y}|$

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



Advantages

① Robust to outliers



Disadvantages

① Convergence usually takes time in MAE

③ Huber Loss

① MSE ② MAE

MSE $\Uparrow$ No outlier $\Uparrow$ Hyperparameter $\Uparrow$

Cost fn = $\begin{cases} \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 & \text{if } |y - \hat{y}| \le \boxed{\delta} \end{cases}$

$$\delta |y-\hat{y}| - \frac{1}{2}\delta^2 \text{, otherwise}$$

⇓

MAE

④ RMSE (Root Mean Squared Error)

$$\text{Cost function} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}$$

Advantages

Disadvantages

② Loss or Cost function For Classification Problems

Classification ⟶ CROSS Entropy
- ⟶ Binary CROSS Entropy {Binary}
- ⟶ Categorical CROSS Entropy {Multiclass}
- ⟶ Sparse Categorical CROSS Entropy ↳ {Multiclass}

① Binary CROSS Entropy

Log Loss

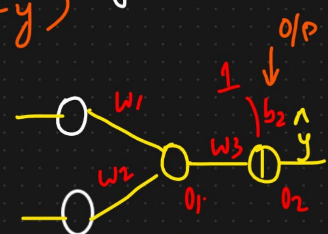
⇓

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

y = Actual value

$\hat{y}$ = Predicted value

Binary classification



Loss = ⇓ Sigmoid

$$\hat{y} \Leftarrow O_2 = \sigma(O_1 * w_3 + b)$$

$$Loss = \begin{cases} -\log(1-\hat{y}) & \text{if } y = 0 \\ -\log(\hat{y}) & \text{if } y = 1 \end{cases}$$

$$\boxed{\hat{y} = \frac{1}{1+e^{-z}}}$$ ⇒ Sigmoid Activation function

② Categorical Cross Entropy (Multiclass classification)

OHE → One Hot Encoding

C = No. of Categories

f_1	f_2	f_3	O/P	Good ($j=1$)	Bad ($j=2$)	Neutral ($j=3$)
→ 2	3	4	Good	[1	0	0]
→ 5	6	7	Bad	0	1	0
→ 8	9	10	Neutral	0	0	1

$i = 1$ to n

$$\boxed{L(x_i, y_i) = -\sum_{j=1}^{C} y_{ij} * \ln(\hat{y}_{ij})}$$

Actual Value ⇐ $y_{ij} = [\, y_{11} \;\; y_{12} \;\; y_{13} \; \cdots \; y_{1c} \,]$

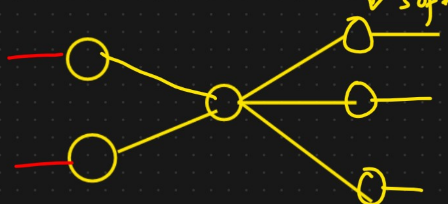
$[\, y_{21}, \; y_{22}, \; y_{23} \; \cdots \; y_{2c} \,]$

$\vdots$

$$y_{ij} = \begin{cases} 1 & \text{if the element is in the class} \\ 0 & \text{Otherwise} \end{cases}$$

Prediction ⇐ $\hat{y}_{ij}$ ⇒ Softmax Activation $= Sof(z) = \boxed{\dfrac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}}$

↙ softmax Activation



O/p $\hat{y}_{ij}$ = Probabilities

[0.1, 0.2, 0.3, 0.2, 0.2] ← 1

Categorical Cross Entropy ⇒ [0.2, 0.3, 0.5]

⇓

[This also gives the probability of other categories]

③ Sparse Categorical Cross Entropy

$$\begin{array}{ccc} 0 & 1^{st} & 2^{nd} \\ \end{array}$$
[0.2, 0.3, $\boxed{0.5}$] → categories

⇓

$\boxed{2^{nd}\text{ Index}}$ ⇒ O/p

$$\begin{array}{ccccc} 0 & 1 & 2^{nd} & 3^{rd} & 4^{th} \end{array}$$
[0.2, 0.3, 0.1, 0.2, 0.2] ⇒ 1^{th} Index

⇓

Category ⇒ O/p

Disadvantage

① Loosing info about the probability of other category.

④ Right Combination

Activation applied

⇓

	Hidden Layers	O/p Layer	Problem Statement	Loss function
①	Relu or its Variants	Sigmoid	Binary Classification	Binary Cross Entropy
②	Relu or its Variants	Softmax	Multi Class	Categorical or Sparse CE
③	Relu or its Variants	Linear	Regression	MSE, MAE, Huber Loss, RMSE