Loss Functions

Regression losses

- Mean Squared error
- Mean Absolute Error
- Huber loss
- Pseudo Huber loss

Classification losses

- Binary Cross Entropy
- Categorical Cross entropy
- Sparse categorical cross entropy
- Hinge loss

Mean Squared error loss.

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

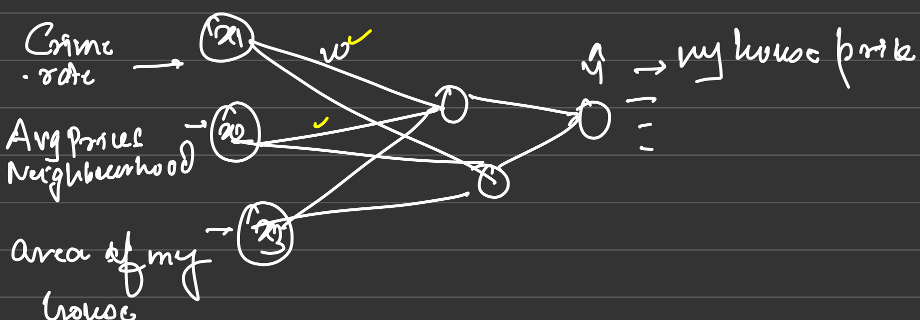
Cost vs Loss

L

x_1	x_2	x_3	$\hat{y}$	y
✓	✓	✓	✓	✓

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

Cost function

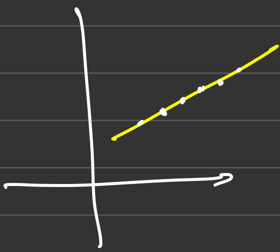


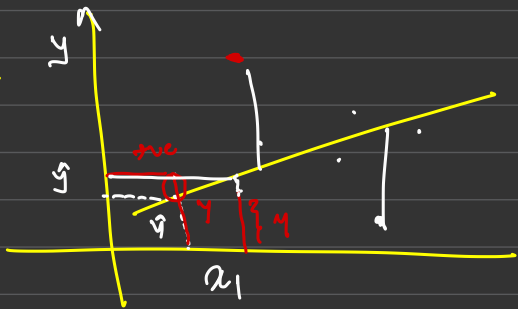
	x_1	x_2	x_3	y	$\hat{y}$
✓	✓	✓	✓	90	70
✓	✓	✓	✓	100	120
✓	✓	✓	✓	110	100

$(y-\hat{y})^2$
400
400
100

$$\text{loss} = \sum_{i=1}^{n} \frac{1}{n} \left(y - \hat{y}\right)^2$$

$$\frac{900}{3} = 300$$





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

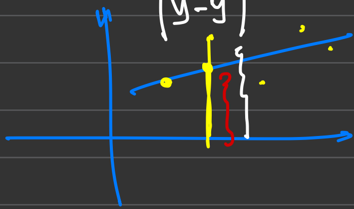
$$(y - \hat{y})$$

MSE is susceptible to outliers

Mean Absolute error

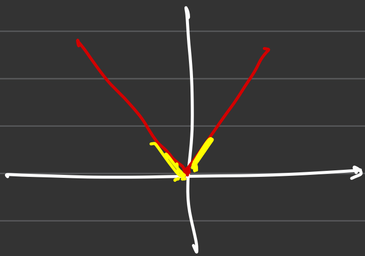
$$\sum_{i=1}^{n} \frac{1}{n} (|y_i - \hat{y}_i|)$$

$|y - \hat{y}|$



x1	x2	y	$\hat{y}$	Loss
✓	✓	20	25	$(\|20-25\|) = 5$
✓	✓	50	72	$(\|50-72\|) = 22$
✓	✓	70	90	$\|70-90\| = 20$

47/3



Huber loss

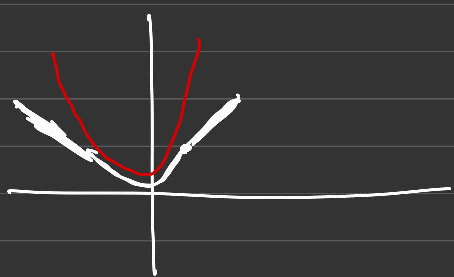
$$= \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{if } |(y_i - \hat{y})| \le \delta \\ \delta\,|y_i - \hat{y}_i| - \frac{1}{2}\delta^2 & \text{if } (|y_i - \hat{y}_i|) > \delta \end{cases}$$

$\delta \rightarrow$ Threshold parameter that determines the point of transition from quadratic to linear loss

Pseudo Huber loss:

$$\delta^2 \left(\sqrt{1 + \frac{(y_i - \hat{y_i})^2}{\delta^2}} - 1 \right)$$

make the Huber loss differentiable



Classification Losses.

Binary Cross Entropy Loss:

x_1	x_2	y	$\hat{y}$
✓	✓	1	~~0.7~~
✓	✓	1	0.6
✓	✓	0	0.1

0.3 ✓ 0.9

$$\text{log loss (BCE)} = -\left[y_i \times \log(\hat{y}_i) + (1-y_i) \times \log(1-\hat{y}_i)\right]$$

$$= 1 \times \log(0.7) + 0 \times \log(0.3)$$

$$= 0.15 ✓$$

$$-\left[1 \times \log(0.3)\right] + 0 \times \cancel{\log(0.7)}$$

$$= 0.52 \uparrow$$

$$- 1 \times \log(0.9) + 0 \times \log(0.1)$$

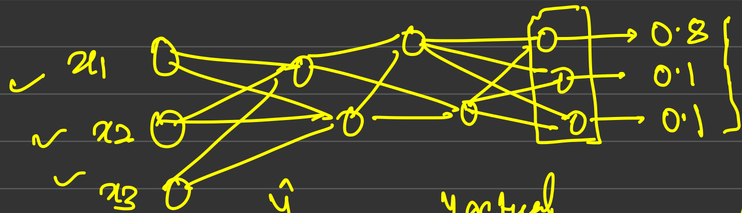
$$= \underline{0.04} ✓$$

$$- y_i \times \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)$$

$$-\left[0 + (1) \times \log(0.9)\right]$$

$$= 0.04$$

Categorical cross entropy
↓
multiclassification problems



$\hat{y}$	y actual			
0.8	1 →	Cat	0	Cat
0.1	0	Dog	1	Dog
0.1	0	Horse	0	Horse

loss b/w $(y, \hat{y})$

$$-\sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij} \times \log \hat{y}_{ij}$$

x_1	x_2	x_3	y			$\hat{y}$		
✓	✓	✓	1	0	0	0.7	0.2	0.1

$$-[1 \times \log(0.7) + 0 \times \log(0.2) + 0 \times \log(0.1)]$$

$$\boxed{\log(0.7)} = 0.154$$

1	0	0	0
0.7	0.2	0.1	0

$y \times \log \hat{y}$

Sparse Categorical CrossEntropy

x_1	x_2	y
✓	~	1
✓	~	2
✓	✓	3

$$\hat{y} \quad \begin{array}{ccc} 1 & 2 & 3 \\ \left[0.7\right. & 0.2 & \left.0.1\right] \end{array}$$

$[0.5 \; 0.4 \; 0.1]$

$[0.7 \; 0.2 \; 0.1]$

$= \log(0.7)$

$[1 \; \cancel{0} \; \cancel{0}]$

$-\log(0.7)$

$-\log(0.4)$

$-\log(0.1)$

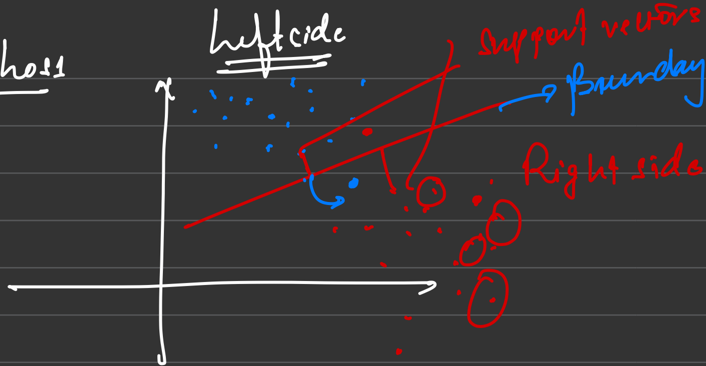
x_1	x_2	y
✓	✓	2

$$\hat{y} \quad \begin{array}{ccc} 1 & 2 & 3 \\ \left[0.6\right. & 0.3 & \left.0.1\right] \end{array}$$

$-\log(0.3)$

$$-\left[0 \times \cancel{\log}\, 0.6 + 1 \times \log(0.3) + 0 \times \cancel{\log}\,(0.1)\right]$$

Hinge loss



$$\max(0\ ,\ 1 - y \times \hat{y})$$

$y \in [-1, 1]$