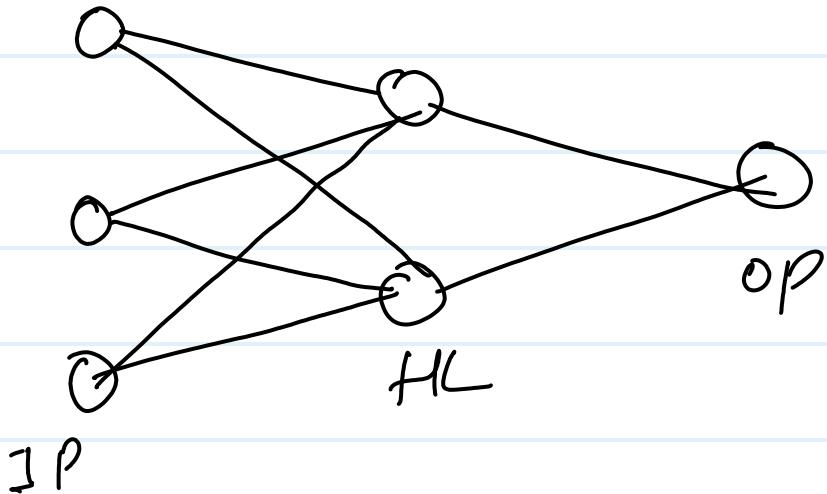


Vanishing Gradient Problem and Activation functions



Activation function

① TanH - zero centric

② ReLU - vanishing gradient solve

③ PreReLU

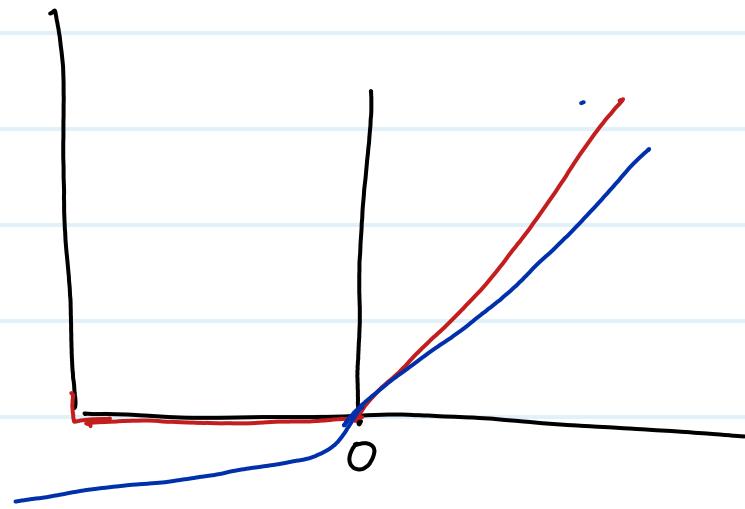
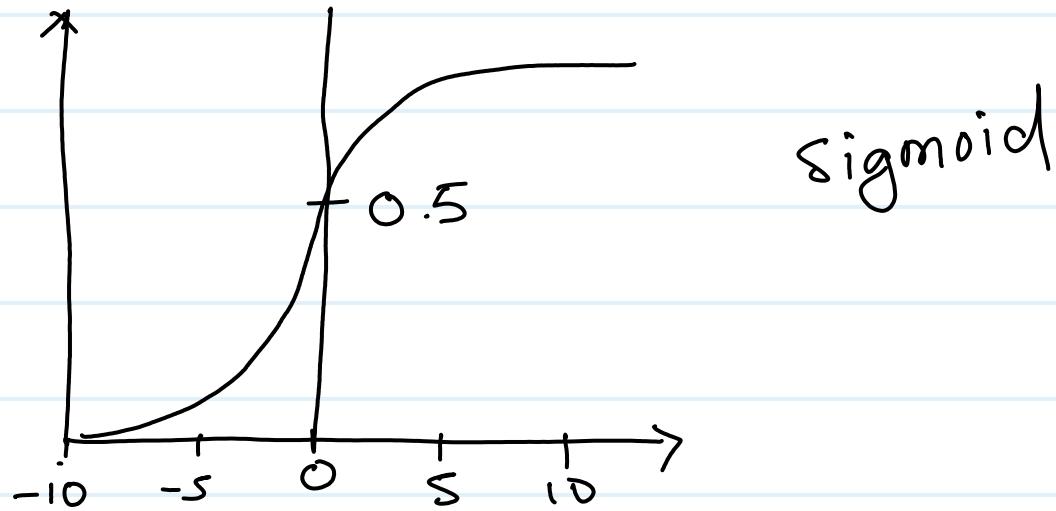
④ ELU Variant of ReLU

⑤ SWISS

⑥ Leaky ReLU

⑦ Sigmoid] \rightarrow Binary class.

⑧ Softmax] O/P \rightarrow multiclass class.



⑨ Linear Activation function

Loss function

Linear Loss fun.

- ① MSE
- ② MAE
- ③ Huber loss

Classification loss fun.

- ① Binary cross entropy
- ② Categorical cross entropy

optimizes

- ① Gradient Decent
- ② SGD (stochastic Gradient Decent)
- ③ mini batch SGD
- ④ SGD with momentum
- ⑤ Adagrad and RMSprop
- ~~⑥ Adam optimizer.~~

Weight Initializing Technique.

- 1) Uniform Distribution
- 2) Xavier / Glorot Initialization
- 3) kaiming He Initialization

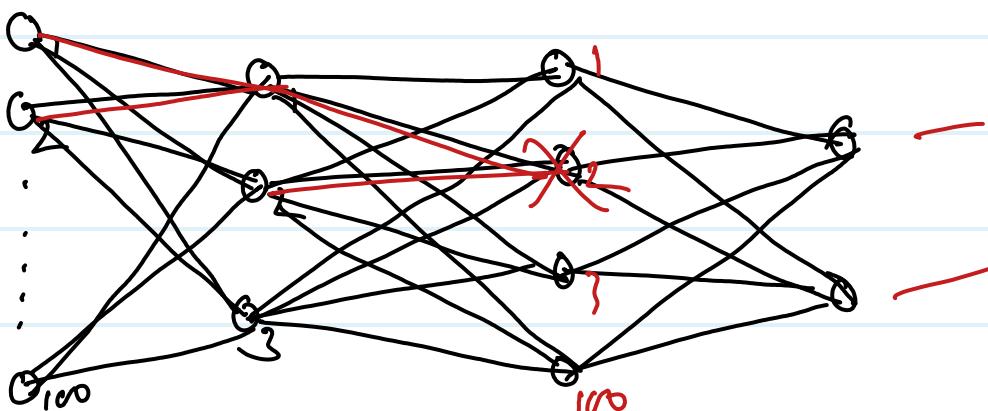
Epoch

$$10000 = 5$$

Batch size \Rightarrow 500

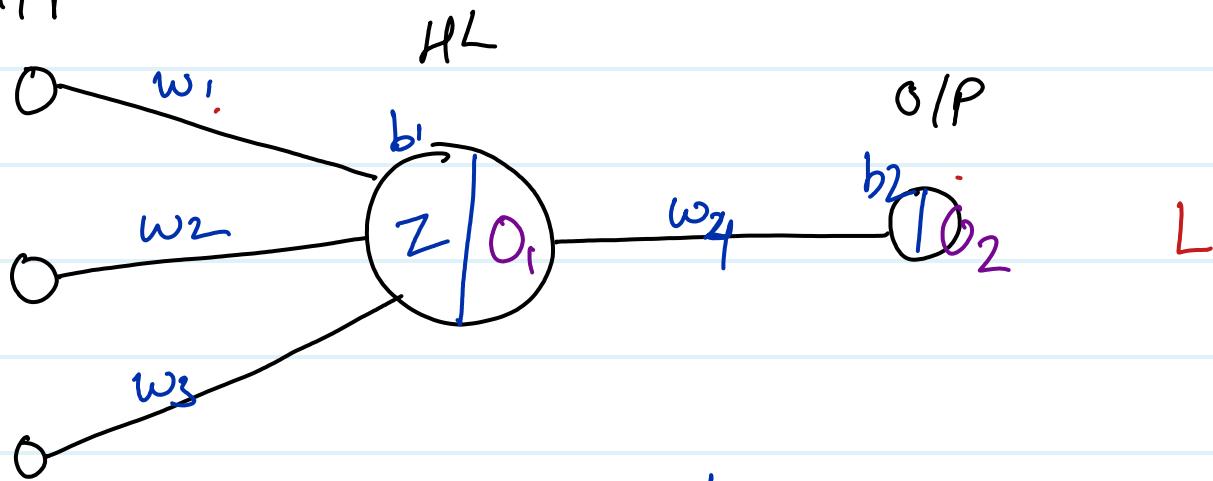
Iteration $\Rightarrow \frac{10000}{500} \Rightarrow 20$

Dropout



x_1	x_2	x_3	y
2	1	4	1
1	3	2	0

I/P



$$Z = xw + b$$

$$\text{Assume } w_1 = 0.01 \quad w_2 = 0.02 \quad w_3 = 0.03 \\ b_1 = 0.001$$

$$Z = 2 \times 0.01 + 1 \times 0.02 + 4 \times 0.03 + 0.001$$

$$Z = 0.161$$

$$\sigma = \frac{1}{1 + e^{-Z}}$$

$$O_1 = \frac{1}{1 + e^{-0.161}} \Rightarrow 0.5460$$

$$w_4 = 0.02$$

$$b_2 = 0.002$$

$$Z = 0.5460 \times 0.02 + 0.002$$

$$Z = 0.0129$$

$$\sigma = \frac{1}{1 + e^{-0.0129}}$$

$$O_2 = 0.503$$

$$\text{Loss} = Y - \hat{Y}$$

$$= 1 - 0.503$$

$$= 0.497$$

optimizer =>

$$w_{4\text{new}} = w_{4\text{old}} - \eta \frac{\partial \text{Loss}}{\partial w_{4\text{old}}}$$

$$b_{2\text{new}} = b_{2\text{old}} - \eta \frac{\partial \text{Loss}}{\partial b_{2\text{old}}}$$

$$\omega_{\text{new}} = \omega_{\text{old}} - \eta \frac{\partial \text{Loss}}{\partial \omega_{\text{old}}}$$

$$\frac{\partial \text{Loss}}{\partial \omega_{\text{old}}} = \frac{\partial L}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial \omega_{\text{old}}}$$