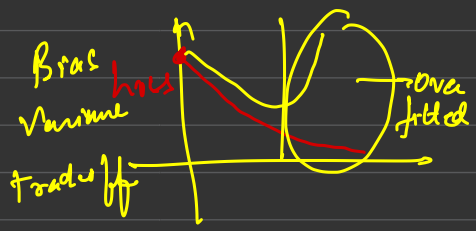
Regularization



less robust (Variance High)

model changes predictions rapidly given slight changes in data

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

(1) $y = [100 + 5x_1 + 2x_2]$ — $\Delta x_1 = 1$ → low variance

$\Delta y = 5$, $\Delta x_2 = 1$, $\Delta y = 2$

$y = 100 + (100 x_1 + 1000 x_2)$ — High variance

if I see a unit change in x_1, 100 units change in y.

" " " x_2 1000 in y

To control / reduce the variance of a model, we can say that β's should not inflate

while training → the abs $|\beta|$ ↑

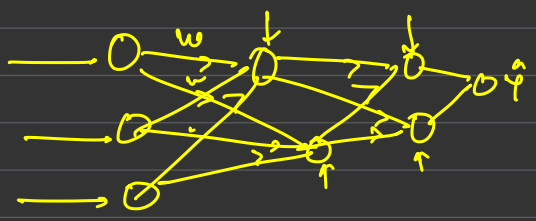
Very high abs $|\beta|$ → high variance

↳ weights and biases

Regularization

- Modifying the loss function
  - Dropout
  - Activation
    - Relu
  - lasso
  - Ridge
  - Elastic Net

Lasso Regression



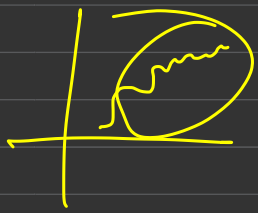
$$\checkmark \quad MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

During your epochs your MSE will reduce

$$\text{new loss} = MSE + \lambda \sum_{i=1}^{p} (|w_i|) \rightarrow \text{Lasso}$$

↓ weight and biases

$$\text{new loss} = MSE + \frac{1}{2} \sum_{i=1} (w_i^2) \rightarrow \text{Ridge}$$



MSE

Lagrangian parameter

①	50 +	②	= 52
②	40	5	= 45
③	32	7	= 39
4	25	11	= 36
5	24	18	= 42

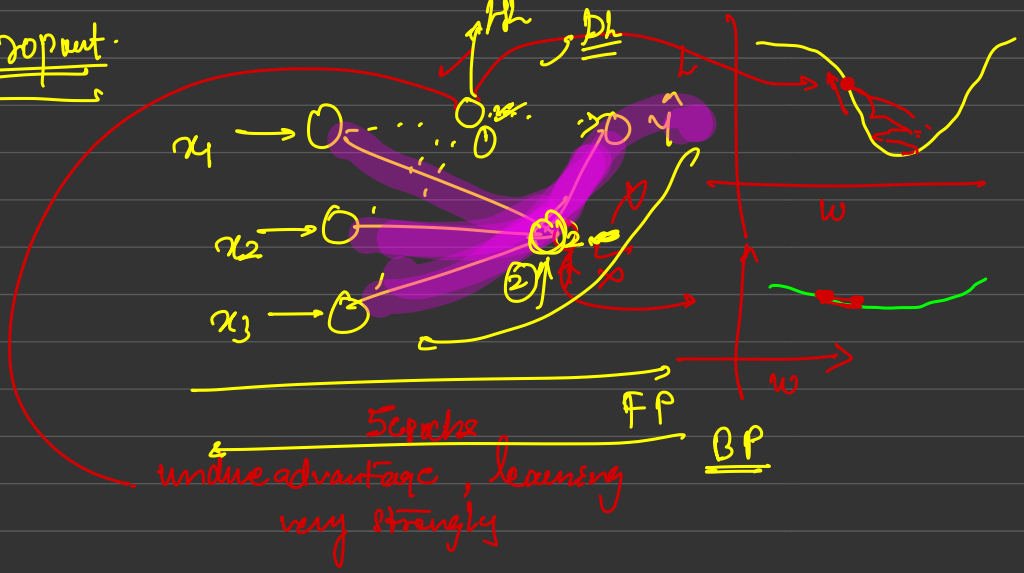
X

Modified loss Elastic Net

$$= \text{MSE} + \lambda\left(\alpha \sum_{j=1}^{p} |B_j|\right) + \left(\frac{1-\alpha}{2} \sum_{j=1}^{p} B_j^2\right)$$

Lasso — Ridge.

Dropout.



Randomly we dropout neurons from the training iteration.

H Neuron will stay (0.8)

(1) T Neuron will dropout (0.2)