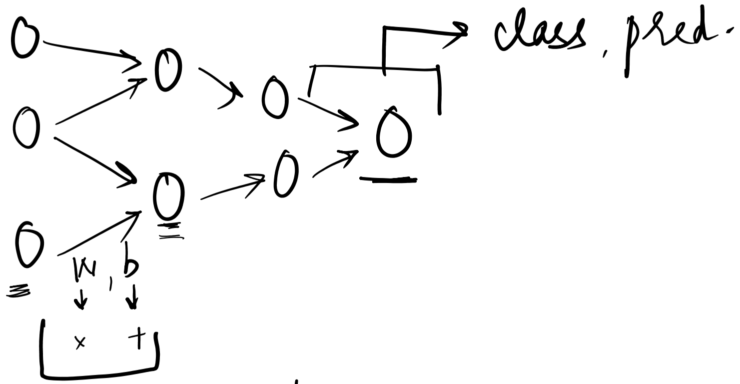
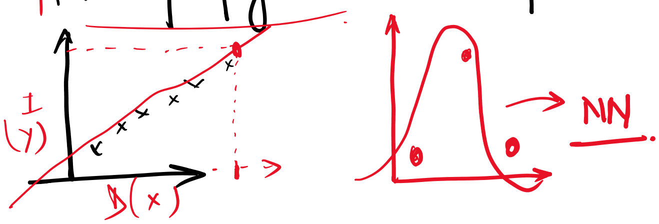
Backpropogation

Friday, May 12, 2023 9:18 PM

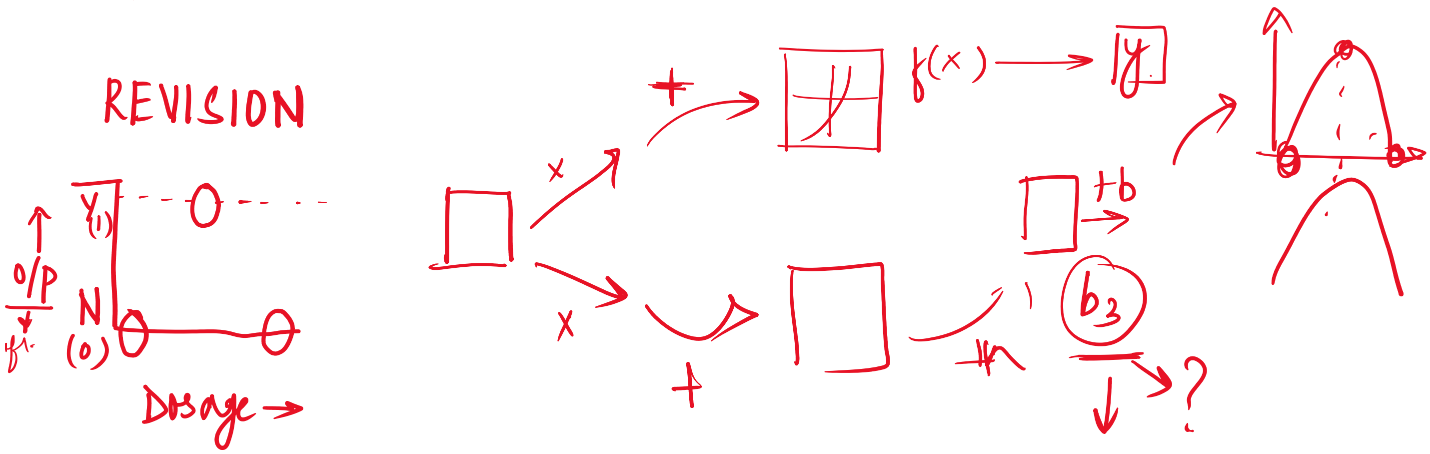


$f(x) = W_1 x + b_1$

* Backpropogation : It helps in optimizing weights & biases



REVISION

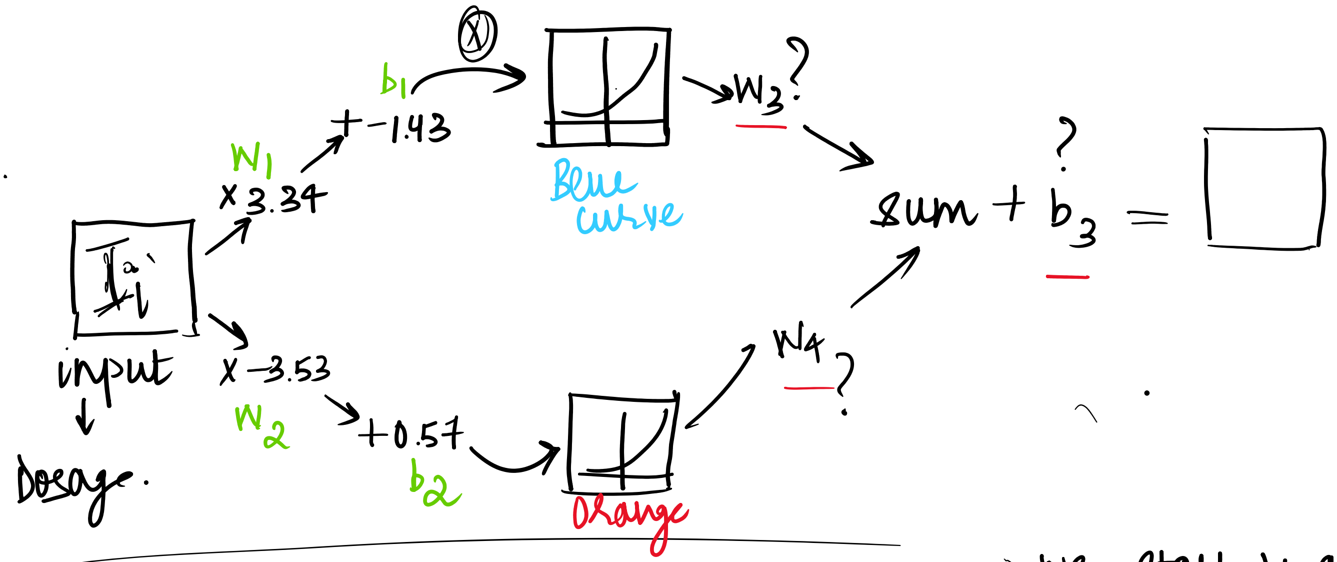


We use the chain rule to find derivative of SSR

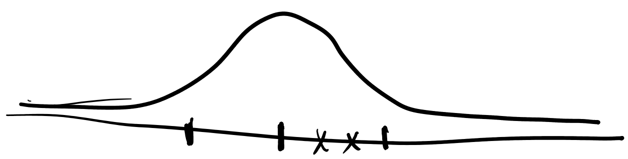
Gradient Descent

Optimal value of b_3

☆ OPTIMIZE 3 PARAMETERS SIMULTANEOUSLY



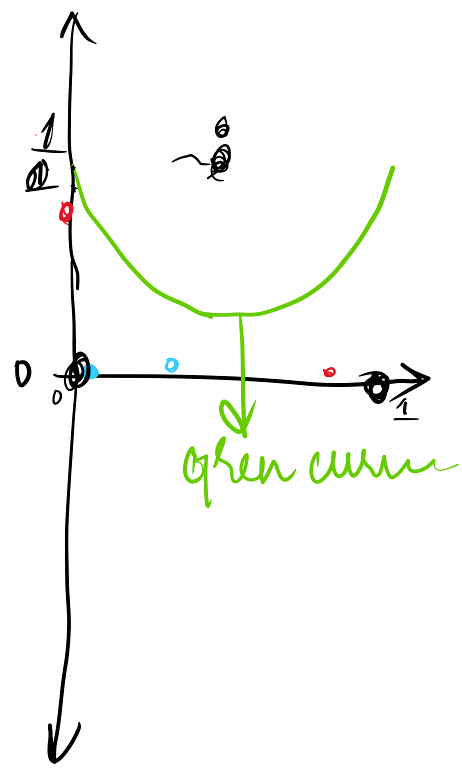
* Standard Normal Distribution



Mean = 0
Std. Dev = 1

We start by assuming
$b_3 = 0$
$W_3 = 0.36$
$W_4 = 0.63$

1. We run Dosage from [0 to 1]
2. Get corresponding y-axis coordinates by using obtained x values in
$f(x) = \log(1+e^x) =$ y-axis coordinates.
↓
Activation fun.

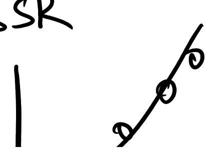


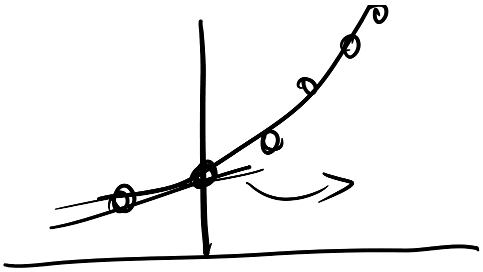
3. Find out how well the green curve fits on the data.

↳ SSR

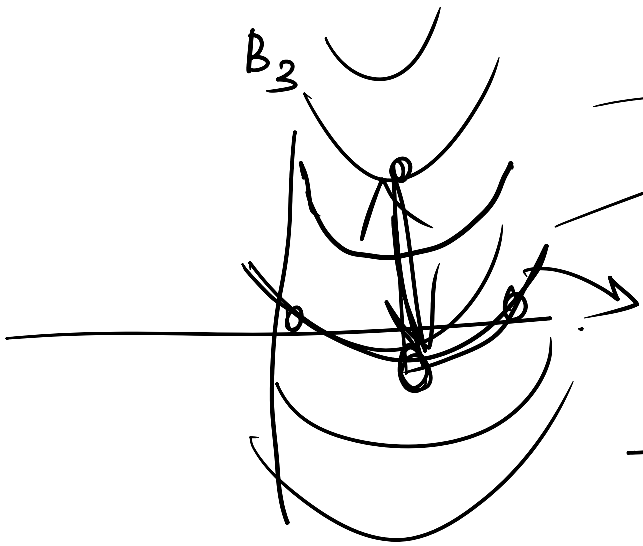
$$SSR = (\text{Obs} - \text{pred})^2 = (0 - 0.72)^2 + (1 - 0.46)^2 + (0 - 0.77)^2$$

SSR





$= \underline{1.4}$



B_3

$\boxed{b_3 = -0.5}$ → this is the optimal value of b_3 when w_3 & w_4 are ass...

$$\frac{dSSR}{db_3} = \frac{dSSR}{dPredicted} \times \frac{dPred}{db_3}$$

↓

$\underline{1.}$

Predicted$_i$ = green − blue
orange
+
b_3

Q How to optimize (how to find derivative of SSR) w.r.t w_3 & w_4.

$$SSR = \sum_{i=1}^{n=3} \left(Observed_i - Pred_i\right)^2$$

$$x_{1,i} = I_i \times 3.34 + -1.43$$

$$x_{1,1} = 0 \times 3.34 - \ldots$$

$$\underline{y_{1,i}} = f(x_{1,i}) = \log(1+e^{x})$$

$*$ Predicted$_i$ = green curve$_i$

$$= \underbrace{y_{1,i} \times w_3}_{\text{blue}} + \underbrace{y_{2,i} \times w_4}_{\text{orange}} + b_3 \; *$$

$$\frac{dSSR}{dw_3} = \frac{dSSR}{d(Pred)} \times \frac{dPred}{dw_3} = \sum_{i=1}^{n=3} -2(O_i - P_i) \times y_{1,i}$$

$$\frac{dSSR}{dw_4} = \frac{dSSR}{d(Pred)} \times \frac{dPred}{dw_4} = \sum_{i=1}^{n=3} -2(O_i - P_i) \times y_{2,i}$$

$$\frac{dSSR}{b_3} = \sum_{i=1}^{n=3} -2 \times (Obs_i - Pred_i) \times 1.$$

$$\frac{dSSR}{dPred} = \frac{d}{dPred} \sum_{i=1}^{n=3} (Obs_i - Pred_i)^2.$$

$$= \sum_{i=1}^{n=3} -2 \,(Obs_i - Pred_i) \quad = -2$$

$$\frac{dPred}{dw_3} = \frac{d}{dw_3}\left[y_{1,i} w_3 + y_{2,i} w_4 + b_3\right]$$

$$= y_{1,i} + 0 + 0 = y_{1,i}$$

$$\frac{d\,Pred}{dw_4} = y_{2,i}$$

$$\frac{dSSR}{dw_3} = \sum_{n=1}^{3} -2\,(o_i - p_i) \times y_{1,i}$$

$$= -2 \times (0 - pred_1) \times y_{1,1} - 2 \times (1 - pred_2) \times y_{1,2}$$
$$- 2 \times (0 - pred_3) \times y_{1,3}$$

$$= 2.58.$$

$$\frac{dSSR}{dw_4} = 1.26 \quad , \quad \frac{dSSR}{db_3} = 1.90$$

Step size = derivative of w_3 × LR (0.1) = 0.258.

New w_3 = old w_3 − step size.

$= \boxed{0.10}$, New w_4 , new b_4.

* We repeat this process until the predictions no longer improve very much.

→ Max no of step size.