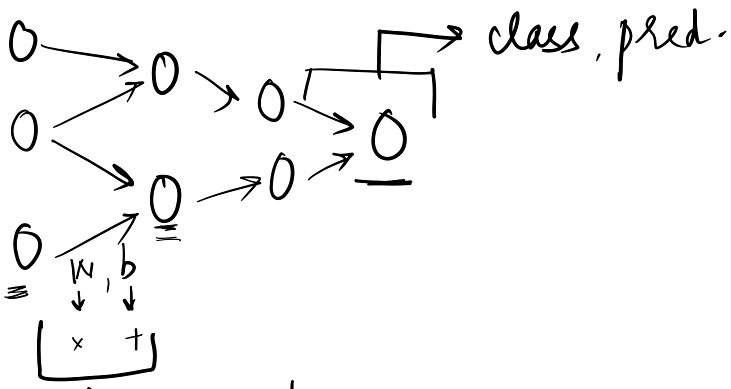
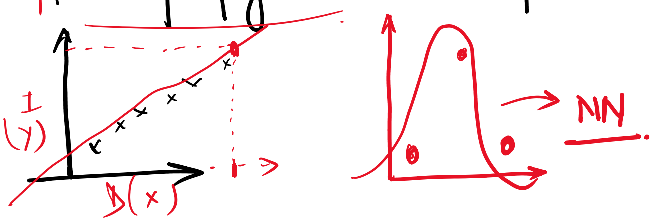
Backpropogation

Friday, May 12, 2023 9:18 PM

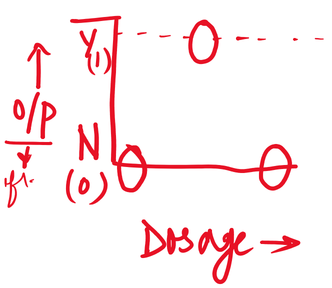


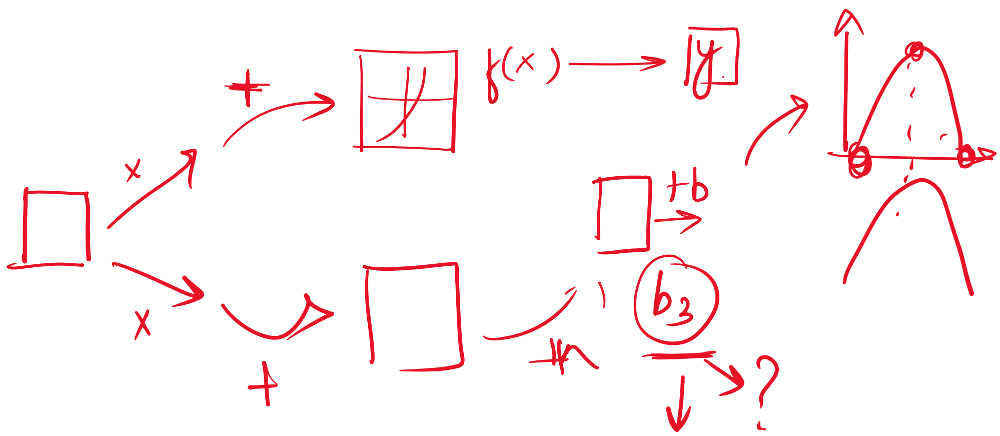
$f(x) = W_1 x + b_1$

* **Backpropogation** : It helps in optimizing weights & biases



REVISION



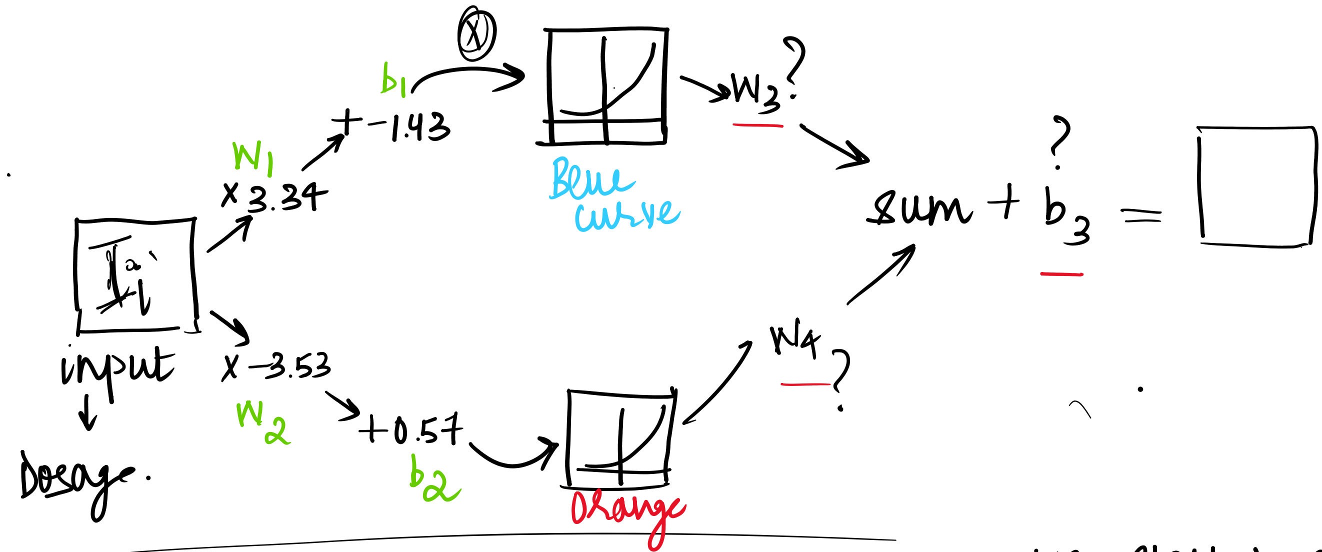


We use the chain rule to find derivative of SSR

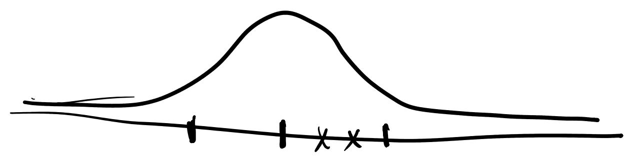
Gradient Descent

Optimal value of b_3

☆ OPTIMIZE 3 PARAMETERS SIMULTANEOUSLY



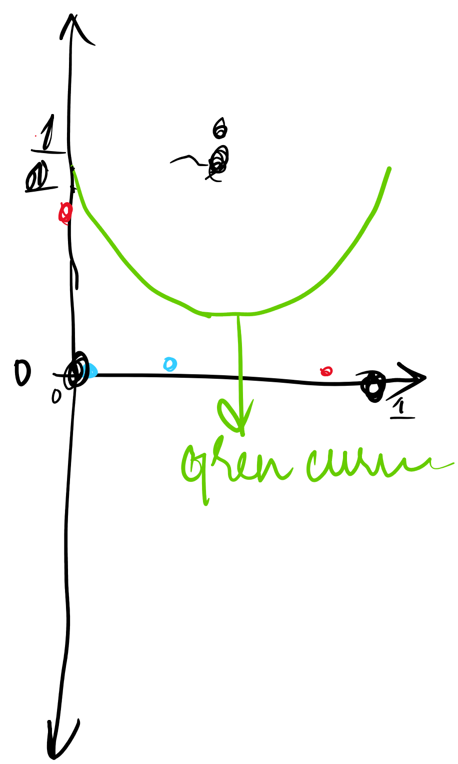
* Standard Normal Distribution



Mean = 0
Std. Dev = 1

We start by assuming
$b_3 = 0$
$w_3 = 0.36$
$w_4 = 0.63$

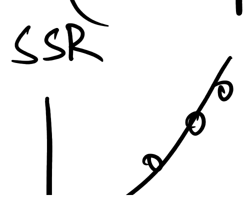
1. We run Dosage from [0 to 1]
2. Get corresponding y-axis coordinates by using obtained x values in
   $f(x) = \log(1+e^x) =$ y-axis coordinates.
   ↓
   Activation fun.

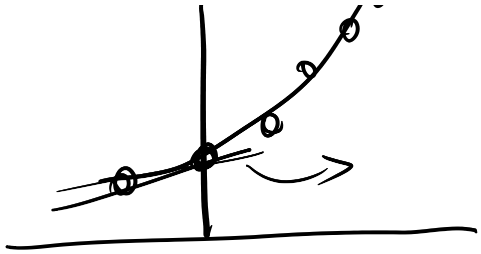


3. Find out how well the green curve fits on the data.

↳ SSR

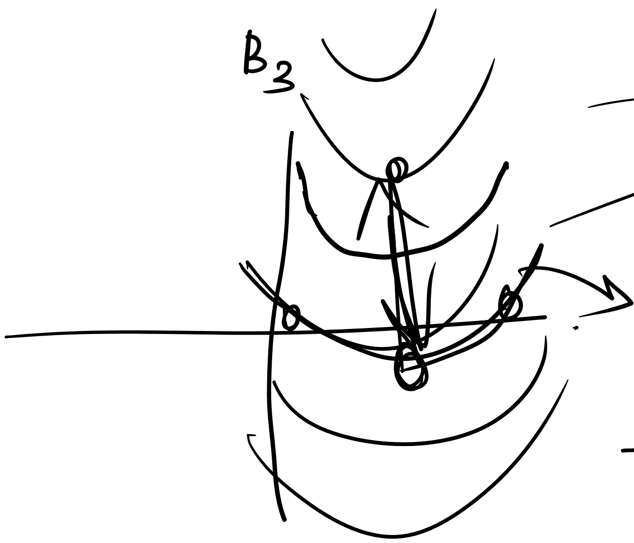
$$SSR = (\text{Obs} - \text{pred})^2 = (0 - 0.72)^2 + (1 - 0.46)^2 + (0 - 0.77)^2$$





$= \underline{1.4}$

$\boxed{b_3 = -0.5}$ → this is the optimal value of b_3 when W_3 & W_4 are ass...



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

$\downarrow$

$\underline{1}$

$Predicted_i$ = green − blue
orange
+
b_3

Q How to optimize (how to find derivatee of SSR) w.r.t W_3 & W_4.

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

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

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

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

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

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

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

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

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

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

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

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

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

$$dPred = y_{2,i}$$

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

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

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

$$= 2.58$$

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

Step Size = derivative of w_3 × LR = 0.258 (LR = 0.1)

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.

OPTIMIZING ALL PARAMETERS SIMULTANEOUSLY

$\frac{dSSR}{}$, $\frac{dSSR}{}$, $\frac{dSSR}{}$

$$\frac{dSSR}{W_3}, \frac{dSSR}{W_4}, \frac{dSSR}{b_3}$$

$$\frac{dSSR}{W_1} = \underbrace{\frac{dSSR}{dPred}} \times \underbrace{\frac{dPred}{dy_1}} \times \frac{dy_1}{dx_1} \times \frac{dx_1}{dw_1}$$

$$\frac{dSSR}{W_2} = \frac{dSSR}{dPred} \times \frac{dPred}{y_2} \times \frac{dy_2}{dx_2} \times \frac{dx_2}{dw_2}.$$

$$\frac{dSSR}{b_1} = \frac{dSSR}{Pred} \times \frac{dPred}{dy_1} \times \frac{dy_1}{dx_1} \times \frac{dx_1}{db_1}$$

$$\frac{dSSR}{b_2}$$

$x_{1,i} = I_i \times w_1 + b_1$; $f(x_{i,1}) = \log(1+e^x) = y_{1,i}$

Blue $= y_{1,i} \times W_3$

$x_{2,i} = I_i \times W_2 + b_2$; $f(x_{2,i,2}) = \log(1+e^x) = y_{2,i}$

Orange $= y_{2,i} \times W_4$.

Green $=$ Blue $+$ Orange $+ b_3$

$Pred_i = y_{1,i}\underline{w_3} + y_{2,i} \times W_4 + b_3$

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

$d(\log z)$ _ 1

$$x_{1,i} = I_i \times w_1 + b_1$$

$$\frac{d}{dz}(\log z) = \frac{1}{z}$$

$$\frac{d}{dx}e^x = e^x$$

$$\frac{dSSR}{dw_1} = \sum_{i=1}^{n=3} -2 \times (O_i - p_i) \times w_3 \times \frac{e^x}{1+e^x} \times I_i$$

$$\frac{dSSR}{dw_2} = \sum_{i=1}^{n=3} -2 \times (O_i - p_i) \times w_4 \times \frac{e^x}{1+e^x} \times I_i$$

$\frac{dSSR}{db_1} =$ As its is $\times$ 1.