Agenda

- → G.D for logistic regression
- → log odds
- → Effect of outliers
- → OVR

GD

→ Repeat until convergence

$$\{ \quad w = W - \eta \frac{\partial L}{\partial W} \quad \}$$

$$\frac{\partial L}{\partial w_d} = -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)} \right) x_d^{(i)}$$

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}} \quad \begin{matrix} = f(x) \\ = g(x) \end{matrix}$$

→ Differentiation of $\sigma(z)$

$$\left[\because \frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x) f'(x) - f(x) g'(x)}{(g(x))^2} \right]$$

$$\frac{d\hat{y}}{dz} = \frac{d}{dz}\sigma(z)$$

$$= \frac{(1+e^{-z}) \cdot 0 - 1 \frac{d}{dz}(1+e^{-z})}{(1+e^{-z})^2} \qquad \frac{d}{dz}(1) = 0, \quad \frac{d}{dz}(e^{-z}) = -e^{-z}$$

$$= \frac{+e^{-z}}{(1+e^{-z})^2} = \underbrace{\frac{1}{(1+e^{-z})}}_{\sigma(z)} \cdot \underbrace{\frac{e^{-z}}{(1+e^{-z})}}_{1-\sigma(z)}$$

$$* \quad \boxed{\frac{d\sigma(z)}{dz} = \sigma(z)\left(1-\sigma(z)\right)}$$

$$1 - \sigma(z) = 1 - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{(1+e^{-z})}$$

$$\hat{y} = \sigma(z)$$

$$\boxed{\frac{d\hat{y}}{dz} = \hat{y}(1-\hat{y})}$$

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

$$\frac{d}{dx}(e^{-x}) = -e^{-x}$$

$$E = -\left[\overbrace{y \log \hat{y}}^{A} + \overbrace{(1-y)\log(1-\hat{y})}^{B}\right]$$

$$\hat{y} = \sigma(z) \qquad z = W^T x + w_0$$

$$E = A + B \qquad = w_1 x_1 + w_2 x_2 - - w_d x_d + w_0$$

$$\frac{\partial E}{\partial w_d} = -\left[\frac{\partial A}{\partial w_d} + \frac{\partial B}{\partial w_d}\right]$$

$$= -\left[\frac{\partial A}{\cancel{\partial \hat{y}}} \cdot \frac{\cancel{\partial \hat{y}}}{\cancel{\partial z}} \cdot \frac{\cancel{\partial z}}{\partial w_d} + \frac{\partial B}{\cancel{\partial \hat{y}}} \frac{\cancel{\partial \hat{y}}}{\cancel{\partial z}} \frac{\cancel{\partial z}}{\partial w_d}\right]$$

$$A = y \log \hat{y} \qquad B = (1-y)\log(1-\hat{y}) \qquad \left[\frac{\partial}{\partial x}(\log x) = \frac{1}{x}\right]$$

$$\frac{\partial A}{\partial \hat{y}} = \frac{y}{\hat{y}} \text{ — } 1 \qquad \frac{\partial B}{\partial \hat{y}} = \frac{(1-y)}{(1-\hat{y})} \cdot \frac{\partial(1-\hat{y})}{\partial \hat{y}} \text{ — } 4$$

$$\hat{y} = \sigma(z) \qquad = \frac{(1-y)}{(1-\hat{y})} \cdot -1$$

$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1-\hat{y}) \text{ — } 2 \qquad d^{th} \text{ feature} \text{ — } 5$$

$$z = w_1 \cdot x_1 + w_2 \cdot x_2 - - w_d \cdot x_d + w_0 \text{ — } 6$$

$$\frac{dz}{dw_d} = x_d \text{ — } 3$$

$$= -\left[\frac{y}{\cancel{\hat{y}}} \cancel{\hat{y}}(1-\hat{y}) x_d - \frac{(1-y)}{\cancel{(1-\hat{y})}} \hat{y}\cancel{(1-\hat{y})} \cdot x_d\right]$$

$$= -\left[y(1-\hat{y}) - (1-y)\hat{y}\right] x_d$$

$$= -\left[y - \cancel{y\hat{y}} - \hat{y} + \cancel{y\hat{y}}\right] x_d$$

$$\boxed{\frac{\partial E}{\partial w_d} = -(y-\hat{y}) x_d}$$

$$\frac{\partial E}{\partial w_1} = -(y-\hat{y}) x_1$$

$$\frac{\partial E}{\partial w_2} = -(y-\hat{y}) x_2$$

$$\frac{\partial E}{\partial w_0} = -(y-\hat{y})$$

$$E = -\left[\overbrace{y \log \hat{y}}^{A} + \overbrace{(1-y) \log(1-\hat{y})}^{B}\right]$$

$$L = \frac{1}{m} \sum E$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log (1-\hat{y}^{(i)}) \right)$$

$$\boxed{\frac{\partial L}{\partial w_d} = -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)} \right) x_d^{(i)}} \rightarrow X^T(y - \hat{y})$$

lin + MSE → convex

logistic + log_loss → Convex

$\hat{y} = w^T x + w_0$

$\hat{y} = \sigma(w^T x + w_0)$

Odds

$$\text{Odd}_{\text{success}} = \frac{\text{No of success}}{\text{No of failure}}$$

$$P_{\text{success}} = \frac{\text{No of succes}}{\text{No succes + No of failures}}$$

$$P_{\text{fail}} = \frac{\text{No of fail}}{}$$

$$\boxed{\text{Odd}_{\text{succ}} = \frac{P_{\text{success}}}{P_{\text{failures}}}} = \boxed{\frac{P_{\text{succ}}}{1 - P_{\text{succ}}}}$$

$$\boxed{P_{\text{suc}} + P_{\text{fail}} = 1}$$

→ Odd of India win against England 4 : 1

$$p = \frac{4}{5}$$

$$q = 1 - p = 1/5$$

$$O = 4/1$$

$$\boxed{\hat{y} = \text{Prob of } y=1 = \frac{1}{(1+e^{-z})} = \sigma(z)}$$

$$\text{Probab of } y=1/x = p$$

$$\text{Prob. } y=0/x = 1-p$$

$$\text{Odds} = \frac{p}{(1-p)}$$

$$\boxed{\text{logits} = \log(\text{odds}) = \log\left(\frac{p}{1-p}\right)}$$

Linear $\rightarrow$ $\hat{y} = w_1 x_1 + w_2 x_2 \,--\, w_d x_d = w^T x + w_0$

logistic $\rightarrow$ Assumption

$$\underbrace{\log\left(\frac{\hat{y}}{1-\hat{y}}\right)} = w_1 x_1 + w_2 x_2 \,--\, w_d x_d$$

$$= w^T x + w_0$$

$$= \boxed{z}$$

$$\frac{\hat{y}}{1-\hat{y}} = e^z$$

$$\frac{1-\hat{y}}{\hat{y}} = \frac{1}{e^z} = e^{-z}$$

$$\frac{1}{\hat{y}} - 1 = e^{-z}$$

$$\frac{1}{\hat{y}} = 1 + e^{-z}$$

$$\boxed{\hat{y} = \frac{1}{1+e^{-z}} = \sigma(z)}$$

$$\log\left(\frac{p}{1-p}\right) = w^T x + w_0$$

$$\left\{ \begin{array}{l} w^T x_1 + w_0 \\ > w^T x_2 + w_0 \end{array} \right.$$

$p_1 = 0.99$

$$\log_e\left(\frac{p_1}{1-p_1}\right) = 4.595$$

$>$

$p_2 = 0.6$

$$\log_e\left(\frac{p_2}{1-p_2}\right) = 0.405$$

$p \uparrow \quad \log(\text{odd}) \uparrow$

$$= \frac{p}{(1-p)}$$

$p > 0.5$

$$\log\left(\frac{p}{1-p}\right)$$

$>0 \longrightarrow p > 1-p$

$<0 \longrightarrow p < 1-p$

$\log_e(1) = 0$

$$\frac{0.5}{1-0.5}$$

→ Break until 22:33 PM

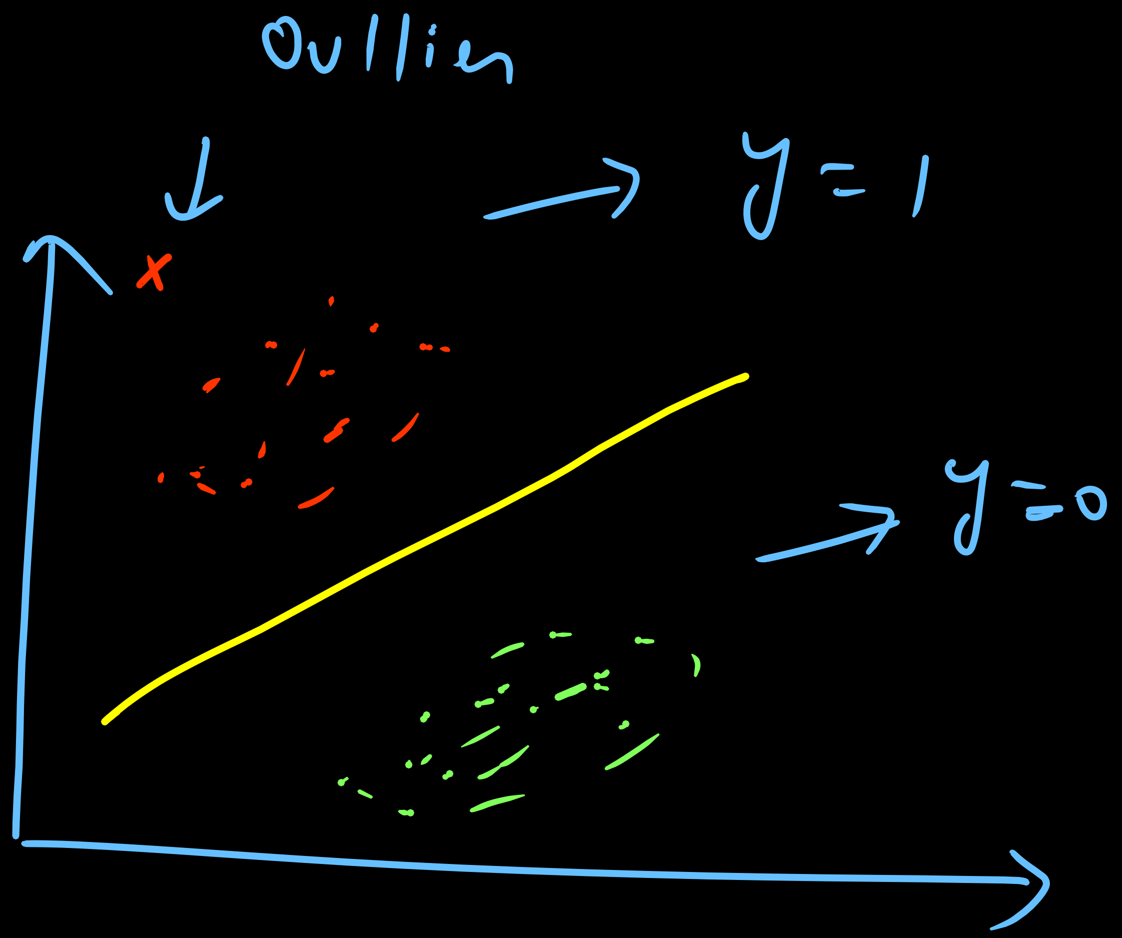
→ Impact of Outlier

1) Outlier is on correct side

$$E = -\left[\overset{1}{y}\log\hat{y} + (1-\overset{1-1}{y})\log(1-\hat{y})\right]$$

$$= -\log\hat{y}$$

$$-\log 0.999 \approx 0 \rightarrow \text{Very low impact}$$



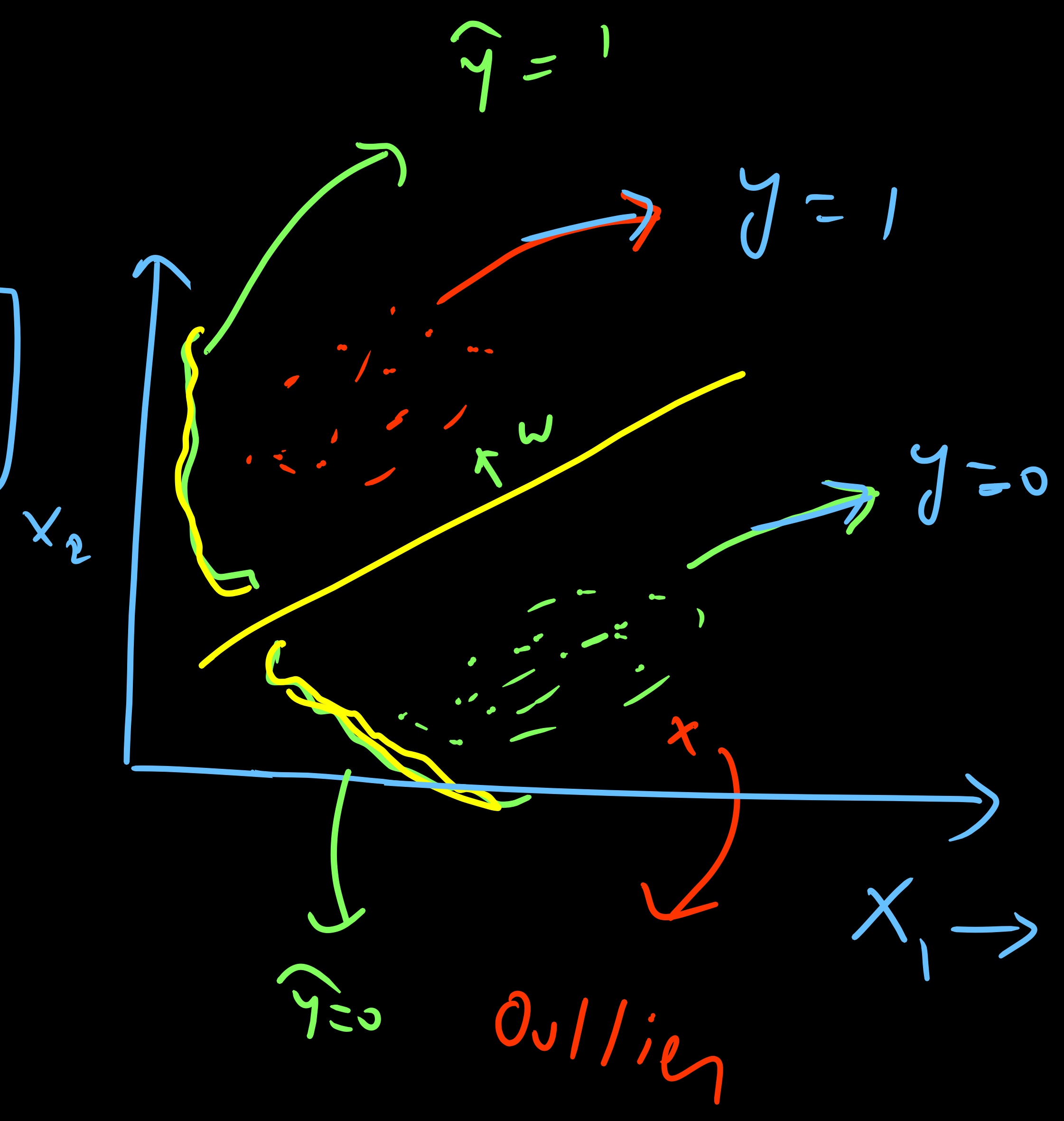
2) Outlier is on incorrect side

$$E = -\left[\underbrace{y \log \hat{y}}_{A} + \underbrace{(1-y)\log(1-\hat{y})}_{B}\right]$$

($y \to 1$, $1 - y \to 1-1$)

$$= -1 \cdot \log(0.001)$$

$$= -6.9$$



x_1	x_2	y
10	0.1	Red
8	0.2	Green
9	0.4	Green
8.5	0.8	Green

$$L = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)}\log \hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right)$$

$$+ \lambda \sum_{j=1}^{d} w_j^2 \longrightarrow L2$$

$C\downarrow \lambda\uparrow$ Underfit

$$+ \lambda \sum_{j=1}^{d} |w_j| \longrightarrow L1$$

$C\uparrow \lambda\downarrow$ Overfit

$$\boxed{C = \frac{1}{\lambda}}$$

$\lambda = 0.001$

$1000 \rightarrow$ Unfit

$\boxed{\text{2 class} \rightarrow \text{Logistic}}$

$\rightarrow$ Multiclass Classification

3 { Orange
Apple
Grape

M1 : Orange or Not
M2 : Apple or Not
M3 Grape or Not

$$\hat{y} = P(y=1/x) = \sigma(W^T x + w_0)$$

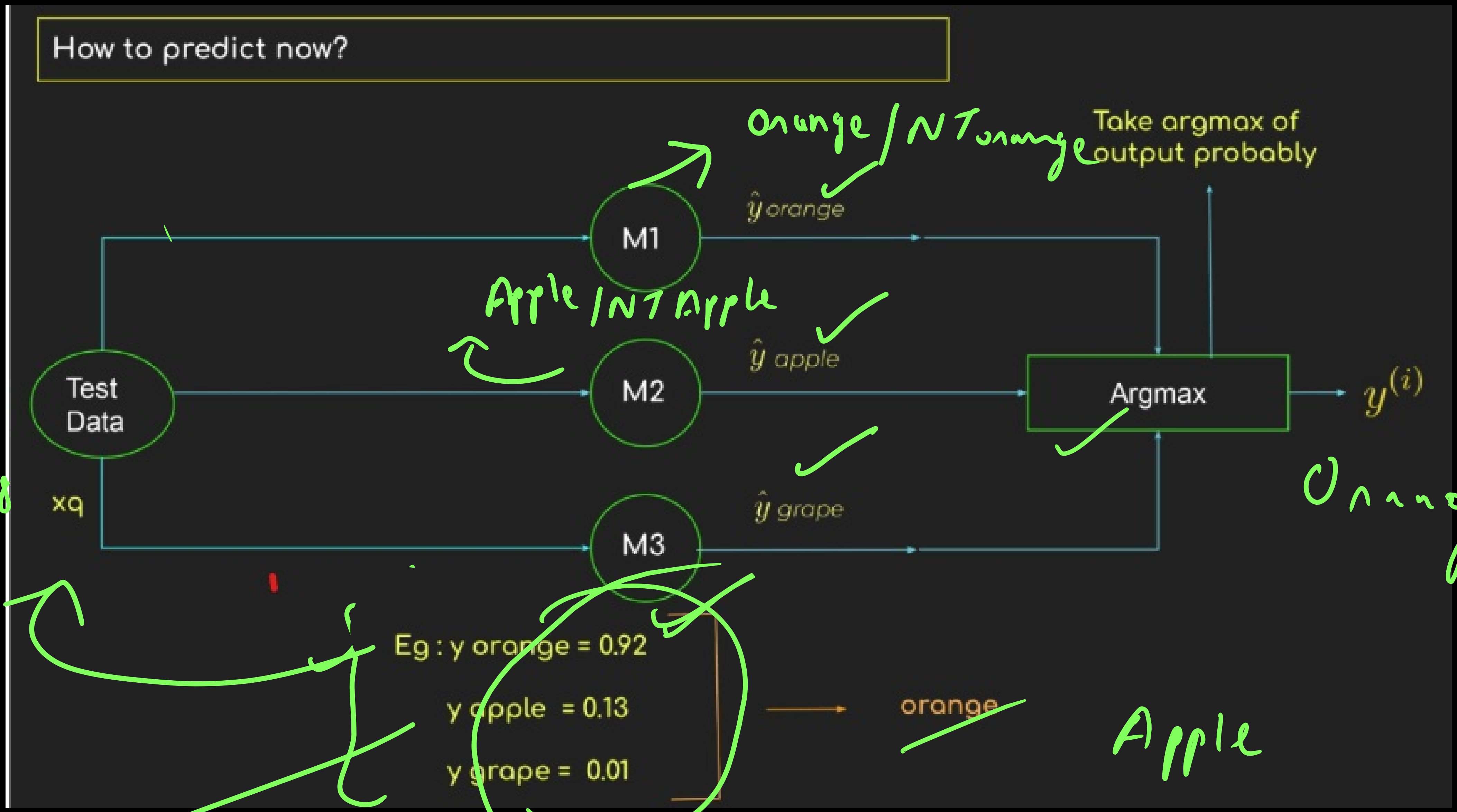
$p = \hat{y}$

$1-p = 1-\hat{y}$

OVR

$P_o = 0.92$

$1-P_o = 0.08$



Orange

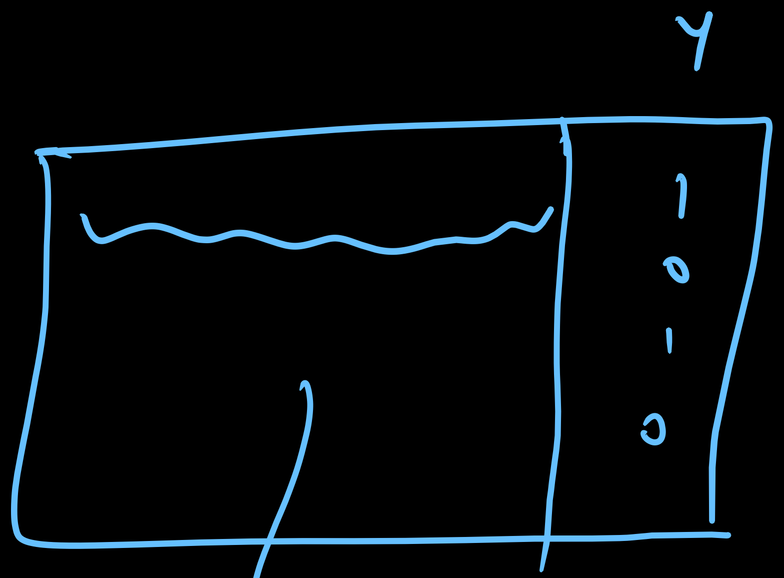
$P_A = 0.13$

$1-P_A = 0.87$

$P_G = 0.01$

$1-P_G = 0.99$

$$\frac{0.92}{0.92+0.13+0.01} = \bigcirc$$



$x_2 = \alpha_1 x_1 + \alpha_3 x_3 \; - - -$

$\longrightarrow$

Standard

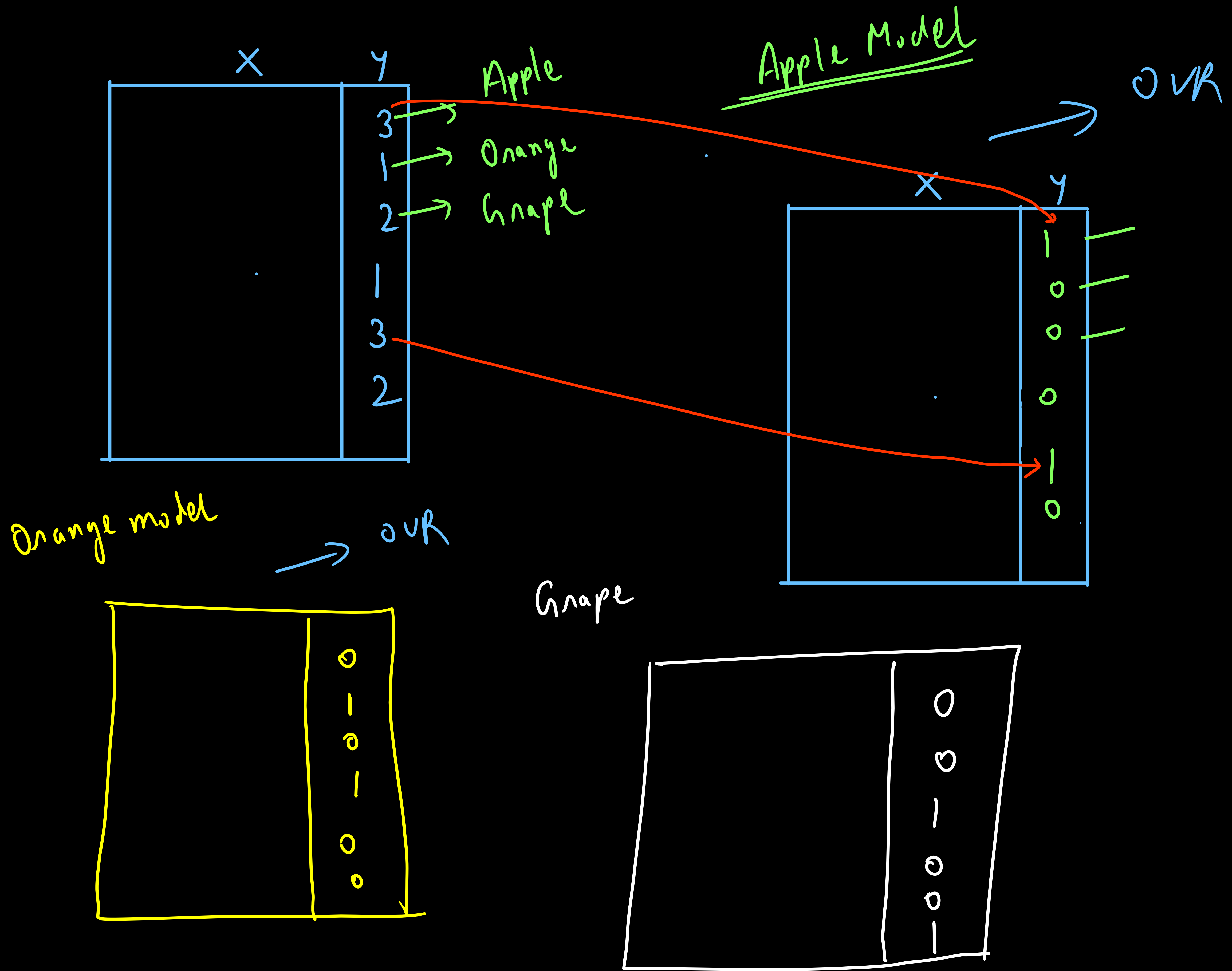
Min Max ✓

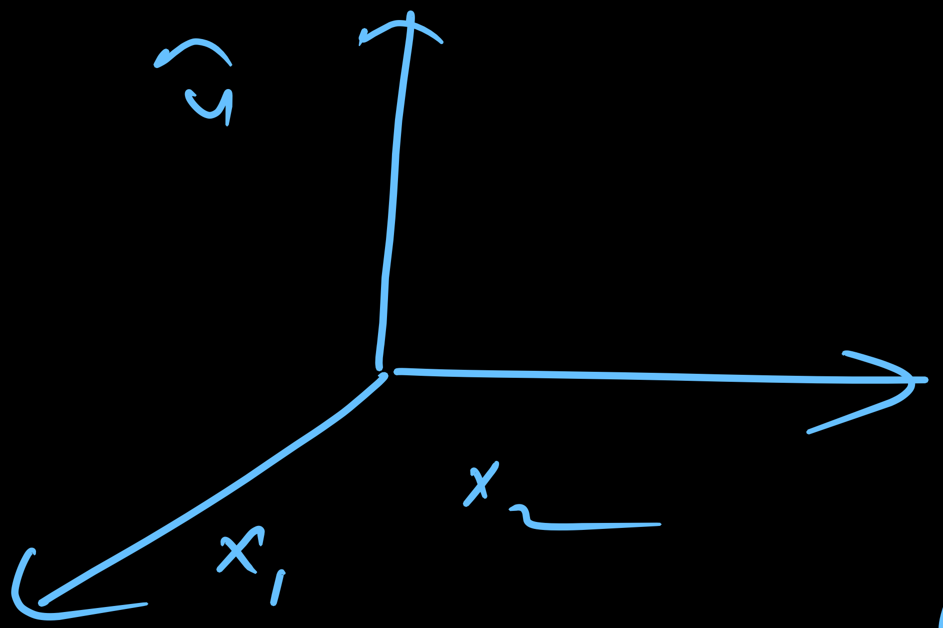
$X_{max} - x$

→ 3 class → 3 Model

→ 1000 Classes → 1000 Model

→ $\boxed{\text{Softmax}}$ → Neural Network


X
Y
3
1
2
1
3
2
Apple
Orange
Grape
Apple Model
OVR
X
Y
1
0
0
0
1
0
Orange model
OVR
0
1
0
1
0
0
Grape
0
0
1
0
0
1



$\hat{y} = w_1 x_1 + w_0$

$n+1$

$= w_1 x_1 + w_2 x_2 + w_3 x_3 + w_0$

4

$w_1 x_1 + w_2 x_2 + w_0$