Logistic-Regression-2

Friday, 21 October 2022 7:37 PM

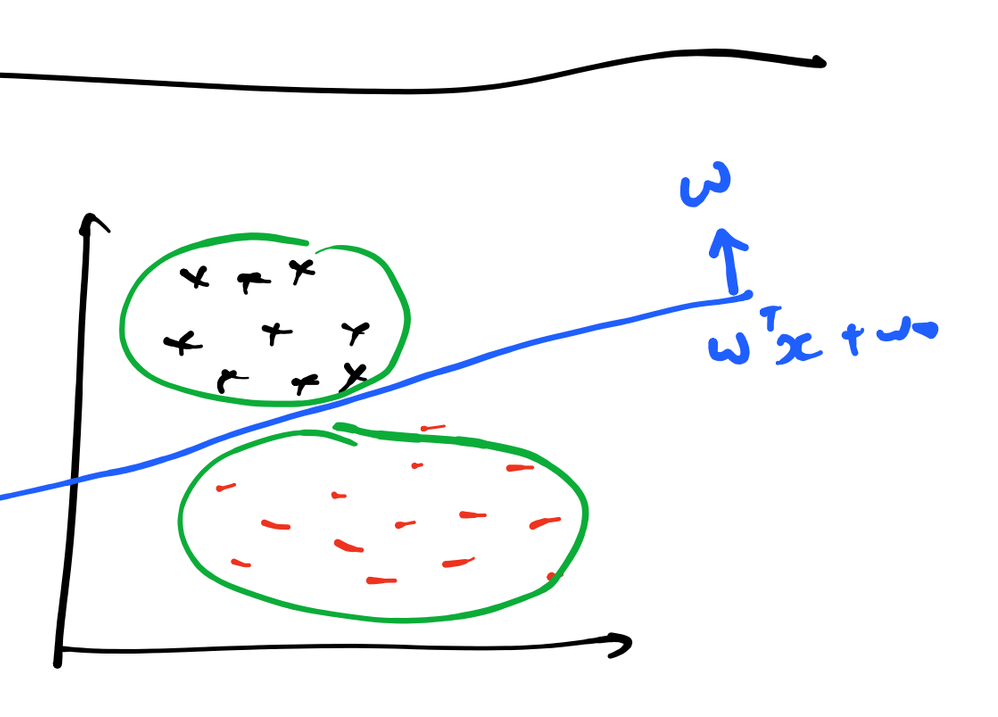
Agenda:

1. Recap
2. Log odds
3. Sklearn Code
4. Multi-class classification
5. Impact of Outliers.

Recap.

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

   Thresholding. $= 0.5$ (typical)



2. $y^{(i)}, \hat{y}^{(i)} \leftrightarrow$ Log-loss

3. G.D — $\frac{\partial L}{\partial w_j}$ — Most.

4. Accuracy : Performance Metric.

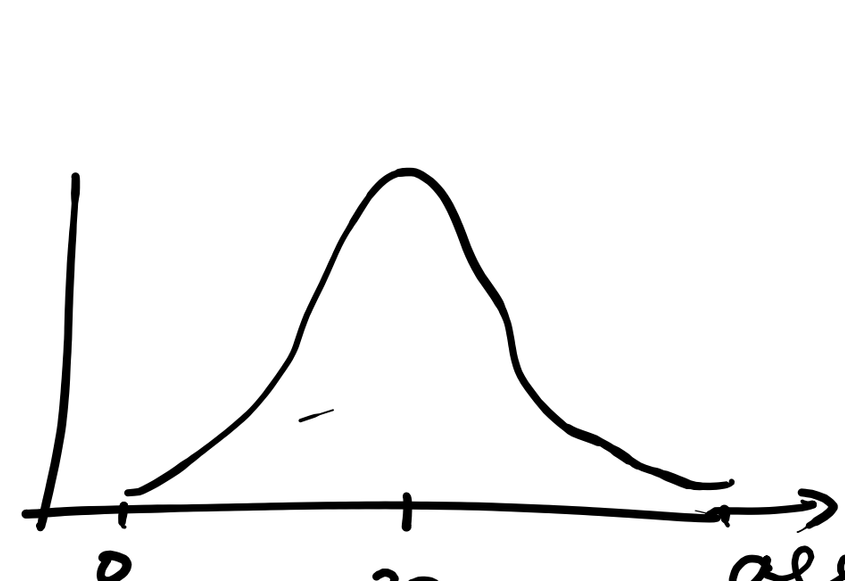
   Not Recommended

← x →	y_{true}	y_{pred}
	1	1
	0	1
	1	0
	1	1
	1	0
	0	0
	0	0

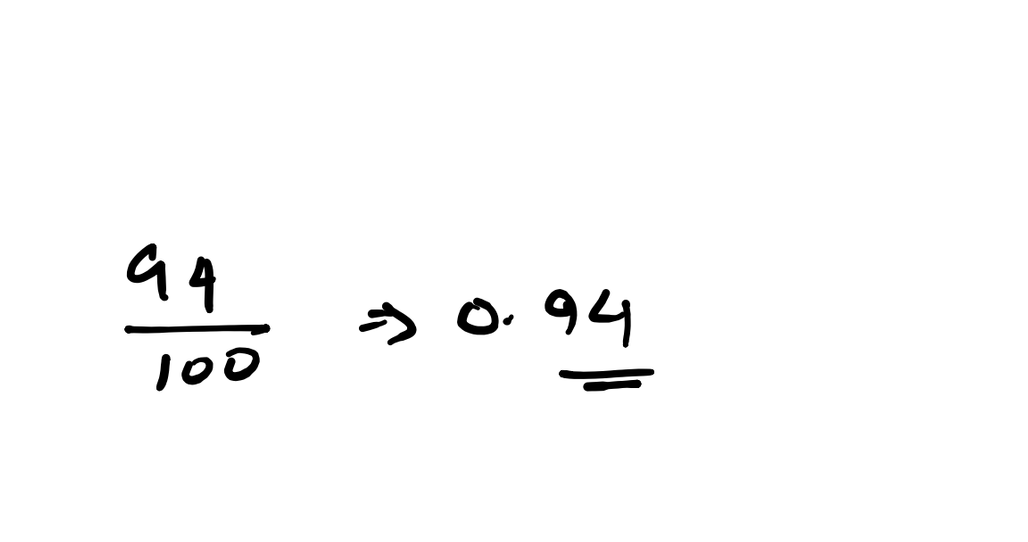
total preds ⇒ 100

Correct pred ⇒ 94

$\frac{94}{100} \Rightarrow 0.94$



Standardization →



Regularisation?

$$\mathcal{L} = \text{log-loss} + \lambda \cdot \sum_{j=1}^{d} w_j^2$$

(reg.)

$$\frac{\partial L}{\partial w_j} \Rightarrow \leftarrow \text{done} \rightarrow \; + 2\lambda \sum_{j=1}^{d} w_j$$

$c \;\; \frac{1}{\lambda}$

$C = \frac{1}{\lambda}$ [C>0]

C ↑ : λ ↓

C ↓ : λ ↑

log odds

English ["odds" : 1 : 4 (W : F)

P(winning) $= \frac{1}{5}$

P(fail) $= \frac{4}{5}$

Imp. $p = P(y=1|x)$ (win, $\hat{y}$)

$1-p = P(y=0|x)$ (fail)

odds $= \frac{p}{1-p}$ = odds of class label being 1

$p = P(y=1|x) = \sigma(w^T x^{(i)} + w_0)$

$$\Rightarrow \frac{1}{1+e^{-(w^T x + w_0)}}$$

$$\frac{1}{1 + \frac{1}{e^{(w^T x + w_0)}}}$$

$a^{-1} = \frac{1}{a}$

$$p \Rightarrow \frac{e^{(w^T x + w_0)}}{e^{(w^T x + w_0)} + 1}$$

$$(1-p) \Rightarrow 1 - \frac{e^{(w^T x + w_0)}}{e^{(w^T x + w_0)} + 1}$$

$$1-p \Rightarrow \frac{1}{e^{(w^T x + w_0)} + 1}$$

$$\text{odds} \Rightarrow \frac{p}{1-p} \Rightarrow e^{w^T x + w_0}$$

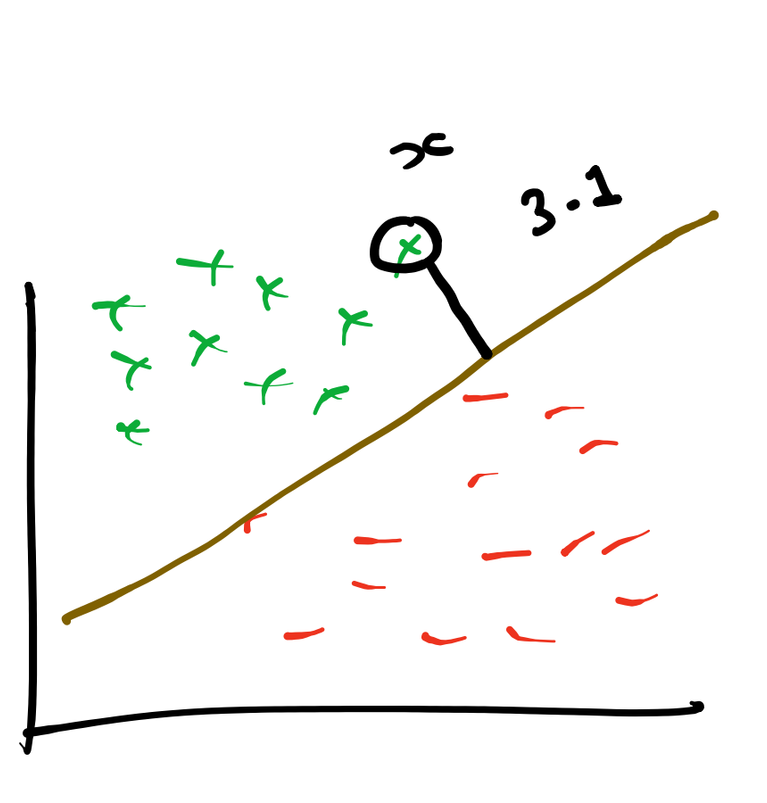
log on both sides.

log(odds): $\log_e\left(\frac{p}{1-p}\right) \Rightarrow w^T x + w_0$ → linear fn. in x

log odds $\log\left(\frac{\hat{y}}{1-\hat{y}}\right)$

$y = [w_0 + w_1 x_1 + w_2 x_2 \dots w_d x_d]$

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

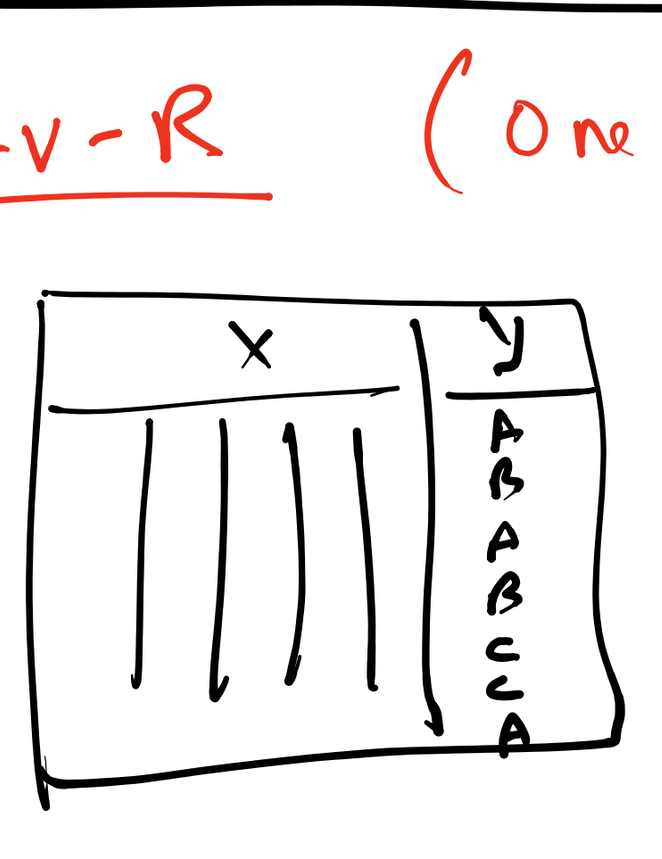


$z = (w^T x + w_0)$

$z = 3.1$

$$\frac{\hat{y}}{1-\hat{y}} = e^{3.1}$$

O-V-R (One Vs. Rest)

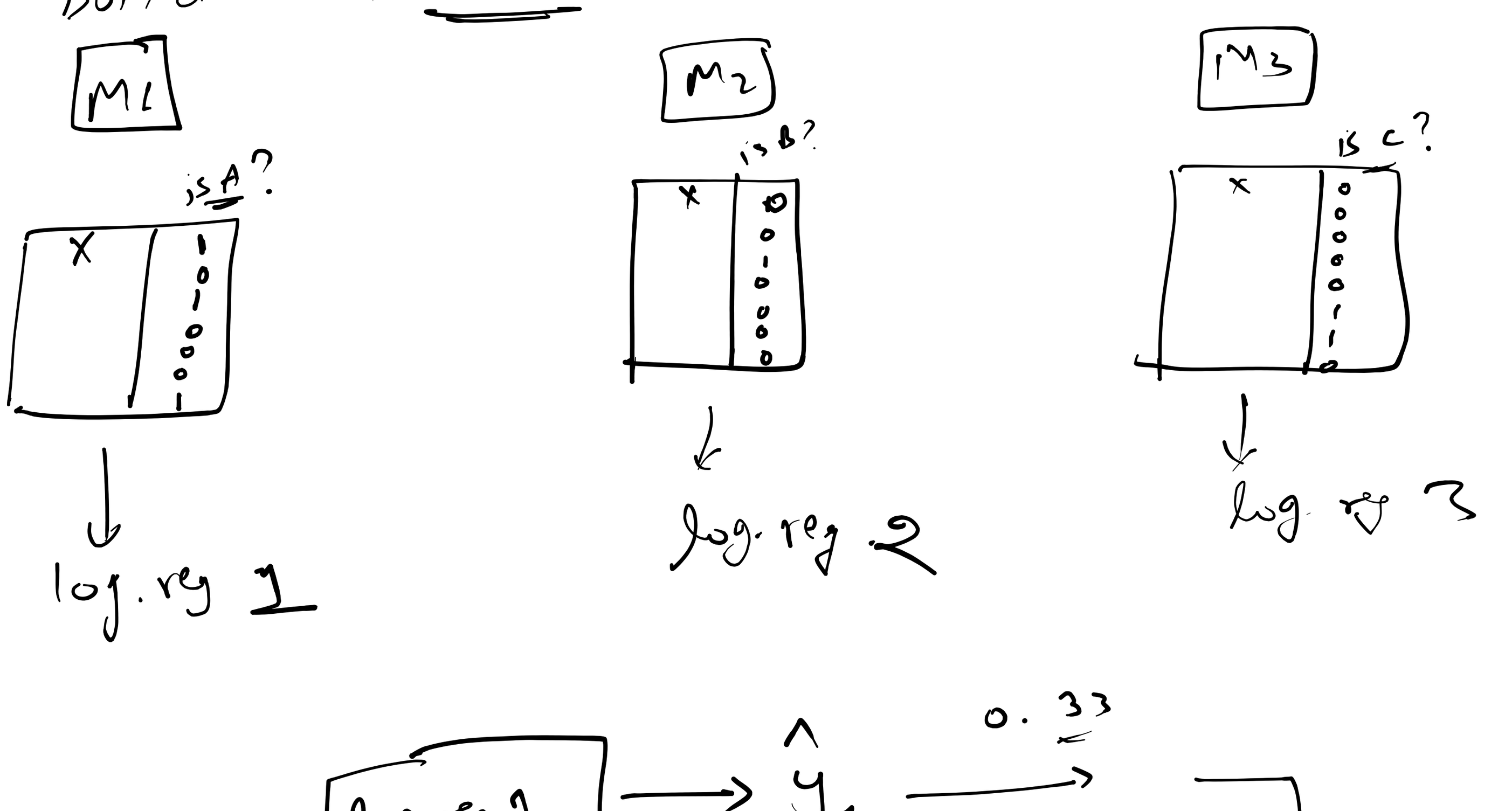


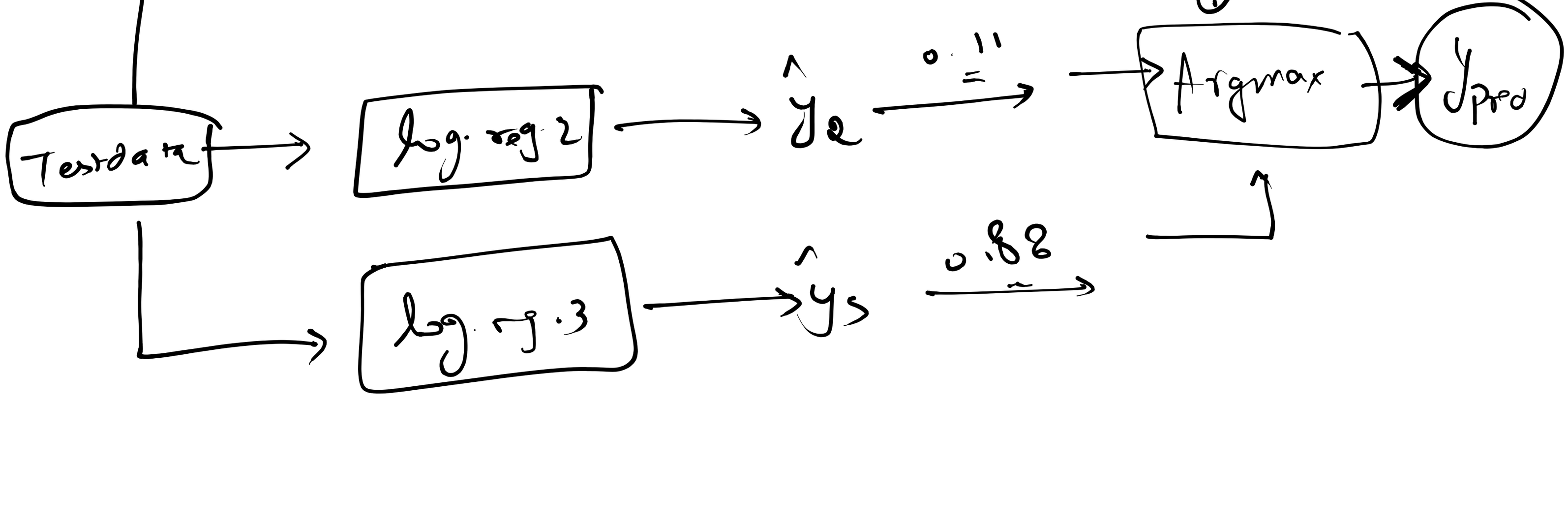
$y \notin \{0, 1\}$

$y \in \{A, B, C\} \dots k$

class 1 : A class 2 : B class 3 : C

Build "3 Models"

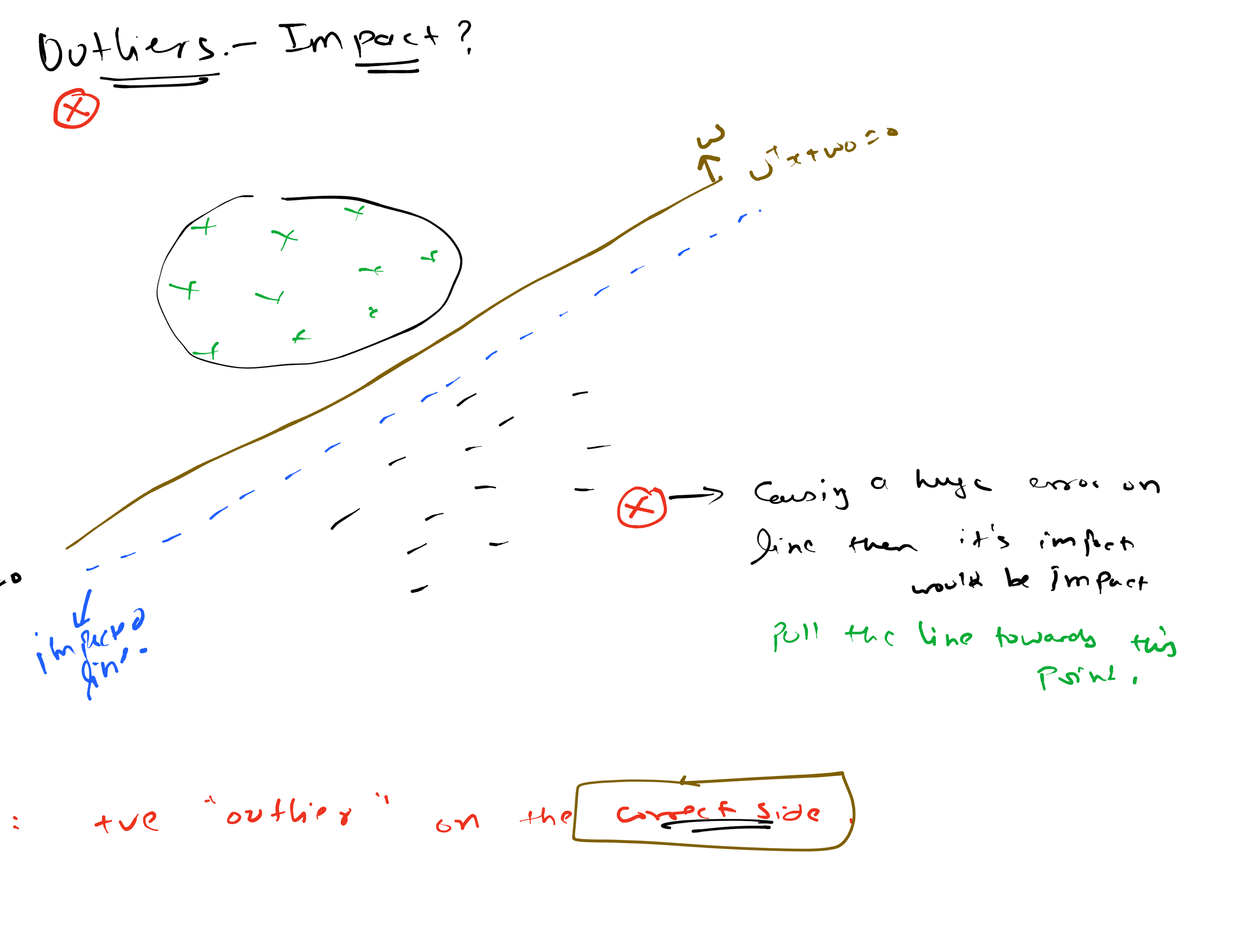




Q → what if there are 1000 of classes?

A → Multinomial log. reg / Softmax

Q → Outliers. – Impact?



⊗ → Causing a huge error on line then it's impact would be impact

Pull the line towards this point.

Case I : +ve "outlier" on the correct side

$w^T x^{(i)} + w_0 \rightarrow$ V. large +ve.

assume $\{\sigma(w^T x^{(i)} + w_0) =$ close to 1 0.999.

$y^{(i)} = 1$.

log-loss $y^{(i)} \Rightarrow 1$

$$\Rightarrow -\log \hat{y}^{(i)} \rightarrow \approx 0.$$

Not much impact on π.

Case II : "Wrong Side"

$y^{(i)} \Rightarrow 1$ loss → $-\log(\hat{y})$ → V. large

$\hat{y} \Rightarrow P(y=1|x) \Rightarrow 1$

0.10