

Logistic Regression 2

- Code
- Accuracy
- Log-odds
- one vs Rest
- Impact of outliers

Accuracy Metric

y	\hat{y}	y_{pred}	
1	0.8	1	✓
0	0.1	0	✓
1	0.4	0	✗

$$\frac{2}{3} = 66.6\%$$

100 test records

⇓

95 are correct

$$\frac{95}{100} \approx 95\%$$

\Rightarrow 1500 samples $\begin{matrix} \nearrow 750 \\ \searrow 750 \end{matrix}$

750 samples for class 1 & class 2

Pred:

$\begin{matrix} 450 - \text{class 1} \\ 350 - \text{class 2} \end{matrix} \quad \left. \vphantom{\begin{matrix} 450 \\ 350 \end{matrix}} \right\} \checkmark \leftarrow$

$$\frac{450 + 350}{1500} = 0.53$$

\Rightarrow Log-odds

odds: 1 : 4

$$p(\text{winning}) = \frac{1}{5}$$

$\overset{n}{\curvearrowright} p = P(y=1 | x)$

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

e

$$\text{odds} = \frac{p}{1-p}$$

$$p = \frac{1}{1 + \exp(-(\mathbf{w}^T \mathbf{x} + w_0))}$$

$$p = \frac{\exp(-(\mathbf{w}^T \mathbf{x} + w_0))}{\exp(-(\mathbf{w}^T \mathbf{x} + w_0)) + 1} = \frac{e^{-2}}{1 + e^{-2}}$$

$$(1-p) = 1 - \frac{\exp(-(\mathbf{w}^T \mathbf{x} + w_0))}{\exp(-(\mathbf{w}^T \mathbf{x} + w_0)) + 1}$$

$$= \frac{1 + \cancel{e^{-2}} - \cancel{e^{-2}}}{1 + e^{-2}}$$

$$= \frac{1}{1 + e^{-2}}$$

$$\left(\frac{p}{1-p} \right) = \frac{\frac{e^{-2}}{1+e^{-2}}}{\frac{1}{1+e^{-2}}} = e^{-2}$$

$$\log\left(\frac{p}{1-p}\right) = -2$$

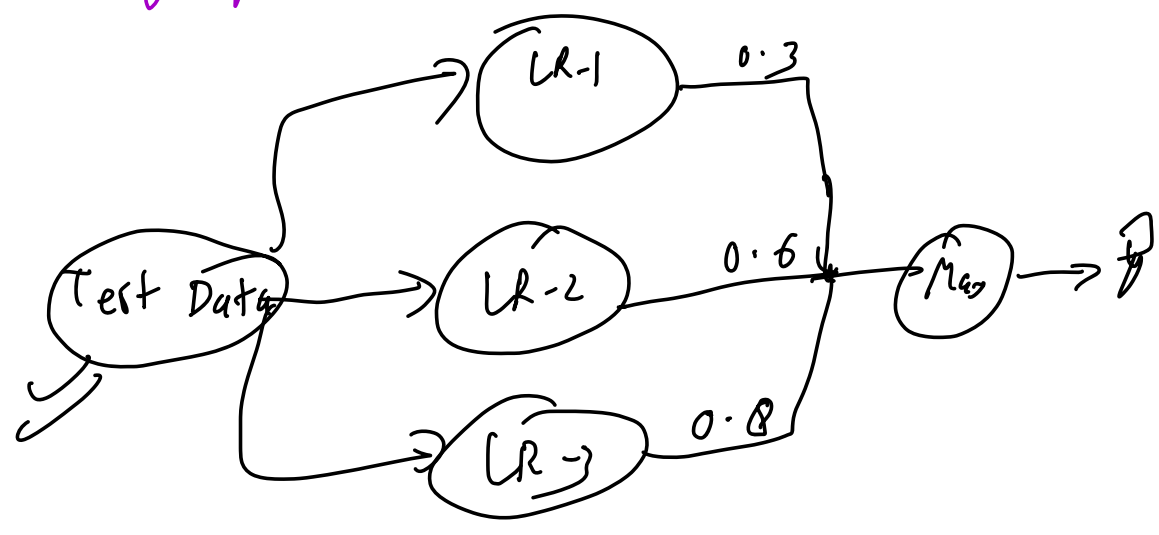
One vs Rest $\Rightarrow > 2$ o/b

f_1	f_2	f_3	y
←		→	Whales 1
			Tuna 0
	↓		Sharks 0
			Whales 1
			Tuna 0
			Whales 1

Log Reg 1 \Rightarrow Whales vs [Tuna, sharks]

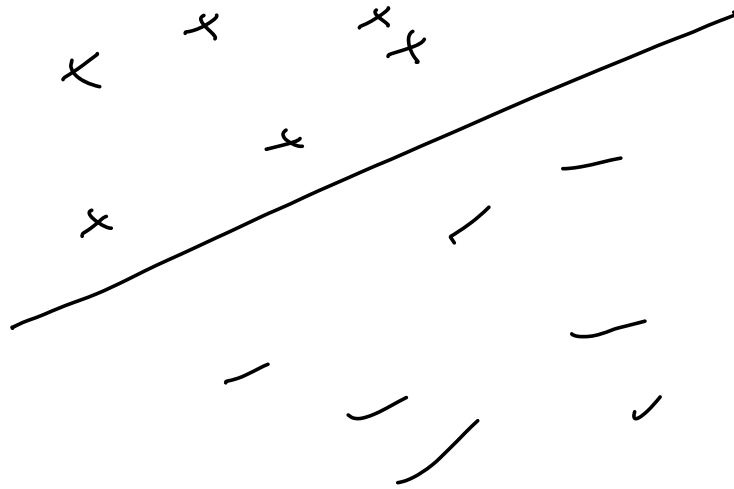
Log Reg 2 \rightarrow Tuna vs [w, I]

Log Reg 3 \rightarrow Shark. vs [w, I]



Impact of outliers

x



$$\begin{aligned} \hat{y} &= 1 \\ \hat{y} &= 0.01 \end{aligned}$$

x