⇒ Recap

⇒ Regularization in Logistic Regression

⇒ Odds interpretation of Hyperplane

⇒ Impact of outliers

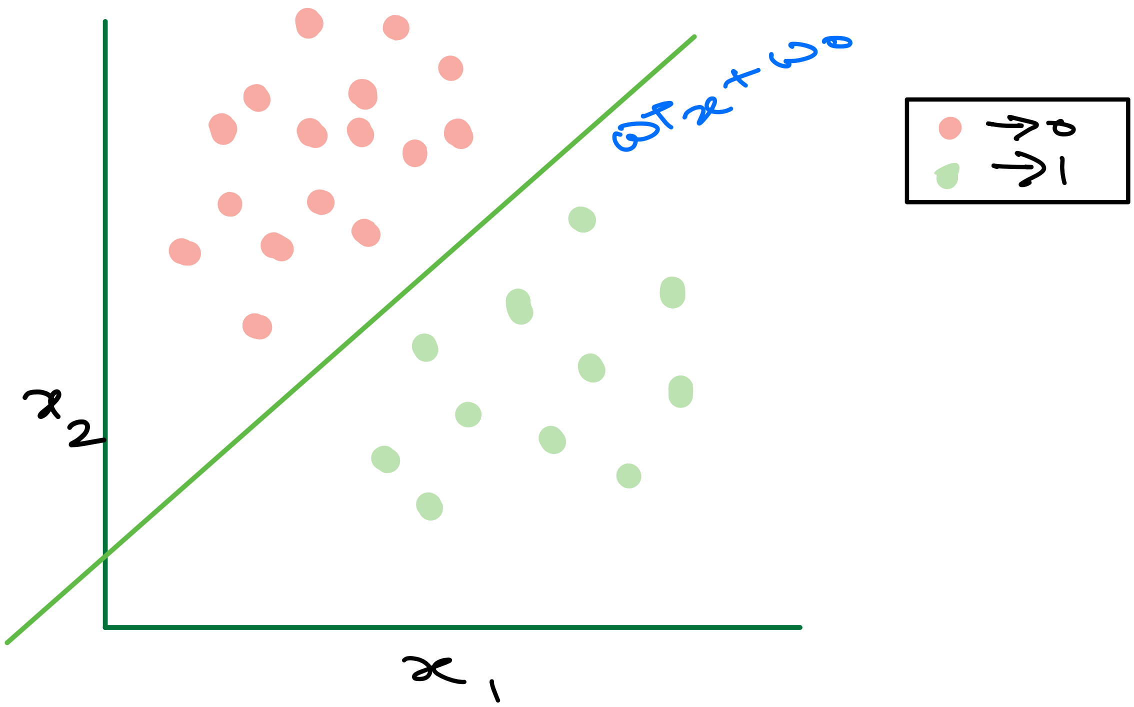
⇒ Multi-Class Classification

Optional

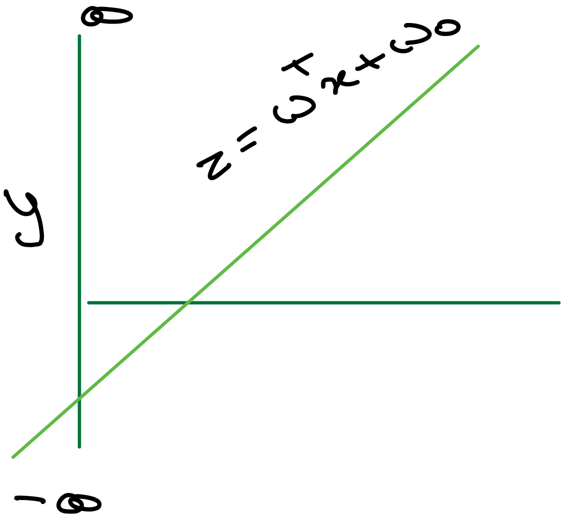
⇒ Probabilities and Likelihood

⇒ Maximum Likelihood Estimation

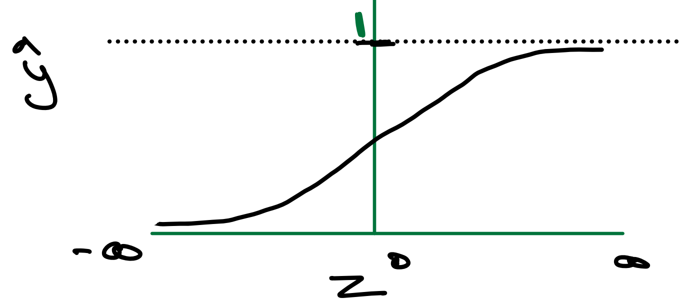
RECAP



$z = w^T x + w_0 \Rightarrow \sigma(z) \to (0, 1)$



$$\sigma(z) \Rightarrow \frac{1}{1+e^{-z}} \Rightarrow \frac{1}{1+e^{-(\omega^T x + \omega_0)}}$$



$$\hat{y}_i = P(y_i = 1)/x_i \Rightarrow \quad 1-P$$

$$\text{Loss}^i \Rightarrow -y^{(i)} \log(\hat{y}^{(i)}) - (1-y^{(i)}) \times \log(1-\hat{y}^{(i)})$$

$$\text{loss} \Rightarrow \frac{1}{n} \sum_{i=1}^{n} -y^{(i)} \log(\hat{y}^{(i)}) - (1-y^{(i)}) \times \log(1-\hat{y}^{(i)})$$

$$\frac{\partial L_A}{\partial \omega_j} \Rightarrow -y(1-\hat{y}) \times x_j$$

$$\frac{\partial L_B}{\partial \omega_j} \Rightarrow (1-y) \times \hat{y} \times x_j$$

$$\text{A} \quad -y(1-\hat{y}) \times x_j$$
$$+$$
$$\text{B} \quad (1-y) \times \hat{y} \times x_j$$

$$(-y + \hat{y} \times y) \times x_j + (\hat{y} - y \times \hat{y}) \times x_j$$

$$x_j(-y + \cancel{y \times \hat{y}} + \hat{y} - \cancel{y\hat{y}})$$

$$x_j(\hat{y} - y)$$

$$\text{Total Loss} \Rightarrow \frac{\partial L_A}{\partial \omega_j} + \frac{\partial L_B}{\partial \omega_j}$$

$$\boxed{\frac{\partial L}{\partial \omega_j} \Rightarrow (\hat{y} - y) \times x_j}$$

Accuracy

threshold = 0.5

x_1 x_2 ... x_d	y	$\hat{y}$	Pred	
—— x_1 ——	1	0.8	1	✓
—— x_2 ——	0	0.3	0	✓
—— x_3 ——	1	0.2	0	✗
—— x_4 ——	0	0.51	1	✗
	1	0.90	1	✓

Compare

$100 \times \left(\frac{3}{5}\right) \Rightarrow 60\%$

Hyperparameter

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

Linear Regress $\Rightarrow$ MSE $+ \lambda w^2$

Logistic Regress $\Rightarrow$ NLL $+ C\, w^2$

$C \Rightarrow \frac{1}{\lambda} \Rightarrow \frac{1}{100} \Rightarrow 0.1$

$0.1 \Rightarrow \sim 100$.

(Higher value of C $\rightarrow$ Less Regularization)

Interpretation of z in terms of log Odd

Odds of Winning

ratio of winning vs Losing

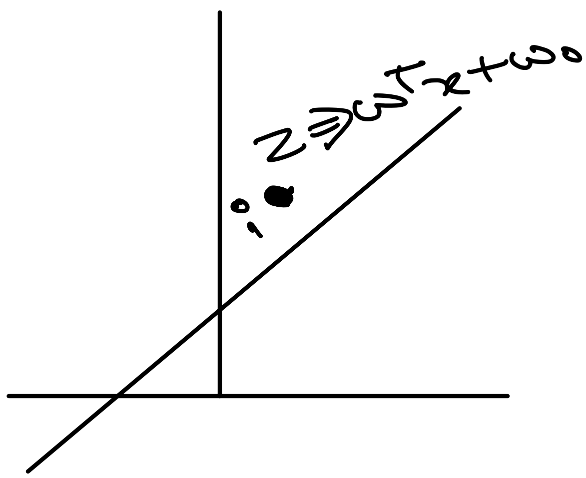
$$\text{Odds} \Rightarrow \frac{12}{1} \qquad 4:1 \Rightarrow \frac{4 \text{ winning}}{1 \text{ Losing}}$$

$$\Rightarrow$$

$$\text{Odds} \Rightarrow \frac{P_{success}}{P_{failure}}$$

$$P_{i=1/x} \Rightarrow \hat{y} \Rightarrow \sigma(z) \quad (i)$$

$$1 - P \Rightarrow 1 - \sigma(z) \quad (ii)$$



$$\boxed{\text{Odds} \Rightarrow \frac{\sigma(z)}{1 - \sigma(z)}} \Rightarrow$$

$$\frac{\log(P)}{\log(1-P)} \Rightarrow z$$

Log Odds Interpretation

$$\text{odd} \Rightarrow \frac{P}{1-P} \Rightarrow \frac{\sigma(z)}{1-\sigma(z)}$$

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

$$1-P \Rightarrow 1 - \frac{1}{1+e^{-z}} \Rightarrow \frac{1+e^{-z}-1}{1+e^{-z}} \Rightarrow \frac{e^{-z}}{1+e^{-z}}$$

$$\text{Odds} \Rightarrow \frac{\frac{1}{\cancel{1+e^{-z}}}}{\frac{e^{-z}}{\cancel{1+e^{-z}}}} \Rightarrow \frac{1}{e^{-z}} \Rightarrow e^{z}$$

$$\text{odds} \Rightarrow e^{z}$$

$$\log_e(\text{odd}) = \log_e(e^{z})$$

$$\log_e(\text{odd}) \Rightarrow z$$

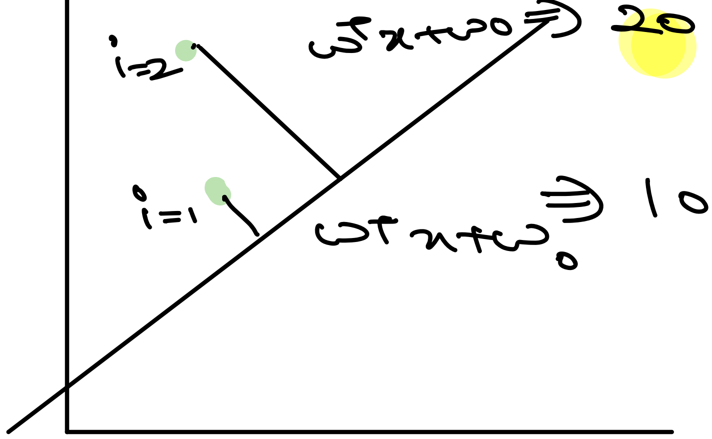
$$\log_e \text{odd} \Rightarrow w^T x + w_0$$

$i=2$

$\log(\text{odds}) \Rightarrow 20$

$i=1$

$\log(\text{odds}) \Rightarrow 10$



The higher the log odds the more will be probability of point belonging to class 1

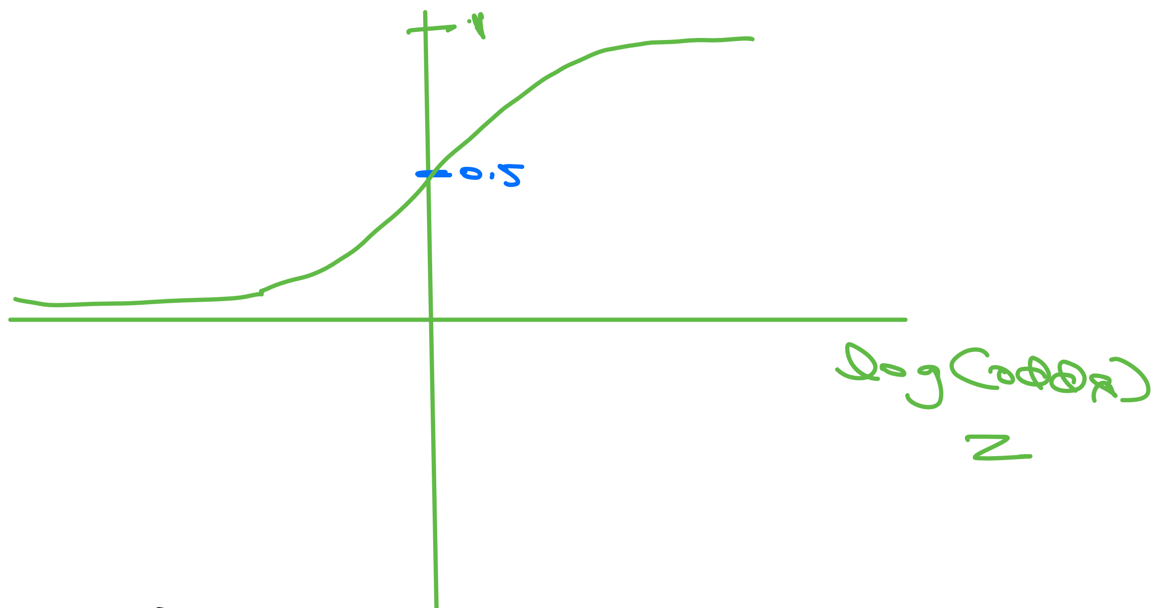
$\log(\text{odds}) \Rightarrow w^T x + w_0 \Rightarrow 100$

$w^T x + w_0 \Rightarrow 5$

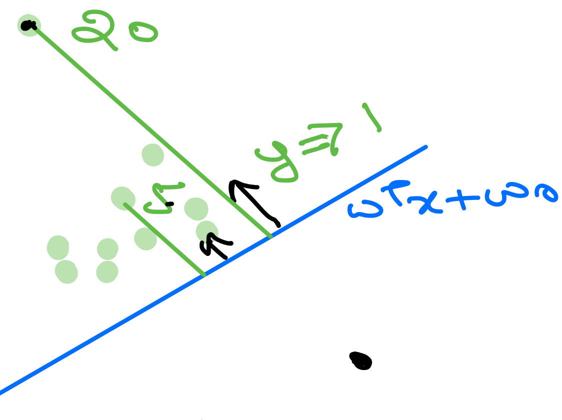
$\log(\text{odds}) \Rightarrow w^T x + w_0$

$\sigma(w^T x + w_0)$

Impact of Outliers on Logistic Regression



Case-1: when the outlier is on correct side



$L \Rightarrow -y_1 \times \log \hat{y}_1 - (1-y_1) \times \log(1-\hat{y}_1)$

$1-1 \Rightarrow 0$

$-y_1 \times \log \hat{y}_1 -$

$w^T x + w_0 \Rightarrow 5$ | $\log(0.993) \Rightarrow 1$ $\quad \frac{1}{1+e^{-5}}$

Not Sign

$w^T x + w_0 \Rightarrow 20$ | $\log(0.9999)\ldots$

1

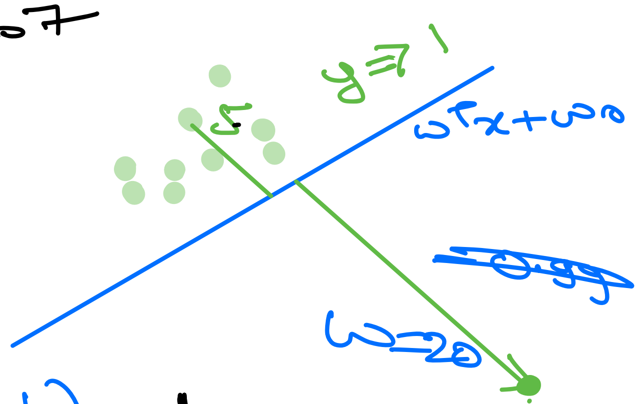
$1 - 0.993$

-0.9993

⇒ Outliers don't have a lot of significance

Case2: When point is not on right Side

(misclassified)



0.007

$w = 5 \Rightarrow \log(0.993)$

$w \Rightarrow -20 \quad \log(0.00001)$

-11.512

Logistic Regression

$\hat{y} \in (0, 1)$

x	y
	1
	2
	3

Tuna
Salmon
Surmayi

3 Clas

$x_1 \longrightarrow \boxed{M_1} \longrightarrow$ Tuna or not

$x_1 \longrightarrow \boxed{M_2} \longrightarrow$ Salmon or Not

$x_1 \longrightarrow \boxed{M_3} \longrightarrow$ Surmaji or not

x	y
———	1
———	2
———	3

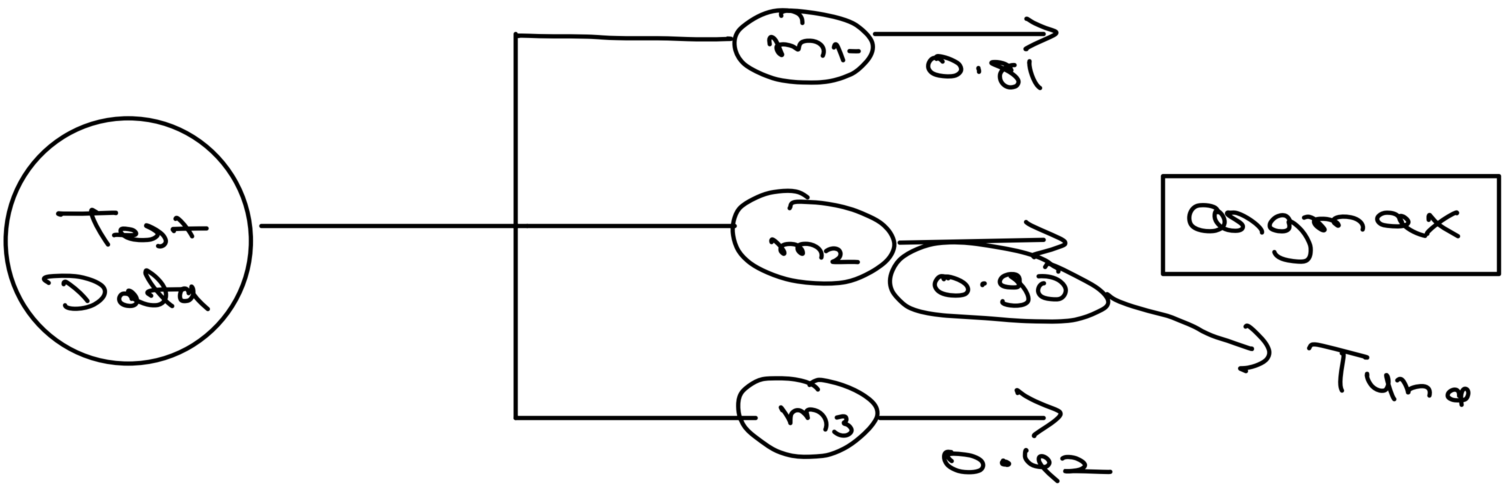
filter all $y_i == 1$ $\rightarrow m_1$

filter all $y_i == 2$ $\rightarrow m_2$

filter all $y_i == 3$ $\rightarrow m_3$

One vs Rest Technique

(OVR)



Doubt Session

StandardSc