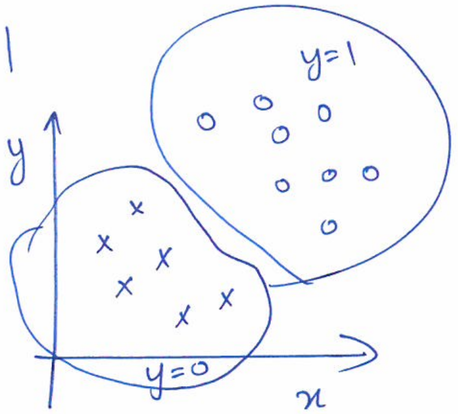
Logistic Regression (Classification)

↓

Binary classifier → 2-class problems.

Logistic Reg ⇒ $0 \le \hat{y}_B(x) \le 1$

- Not a regression
- its a classification (Binary).



Hypothesis fn. for Classification Problem :-

We want $0 \le \hat{y}_B(x) \le 1$

$\hat{y}_B(x)$ → predicted value. B → parameter, x → input features.

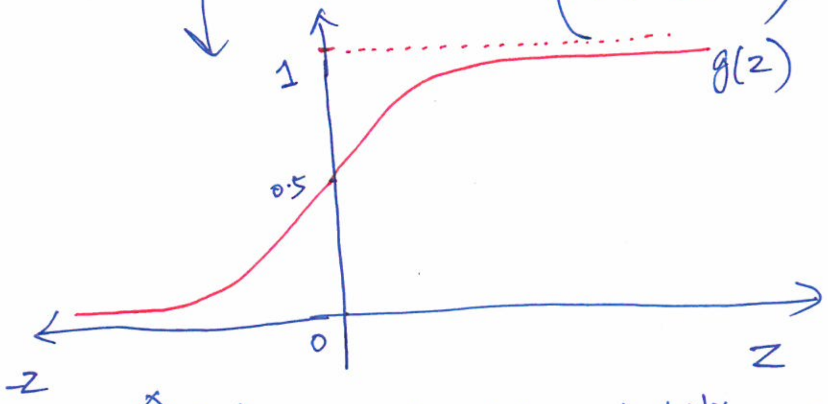
In linear reg : $\hat{y}_B(x) = B^T \cdot X$

For logistic reg: $\hat{y}_B(x) = g(B^T \cdot X)$

where $g(z) = \left(\dfrac{1}{1+e^{-z}}\right)$ → sigmoid fn.

$$\hat{y}_B(x) = \left(\frac{1}{1+e^{-B^T \cdot X}}\right)$$

→ We need to find the parameter 'B' using our training data.



$\hat{y}_B(x)$ = estimated probability that y=1 on input x.

$P(y=1 \mid B, x) = ?$

Example: if $x = [x_0 \quad x_1]^T = [1 \quad \#spam\_word\_count]$

$$\hat{y}_B(x) = 0.7.$$

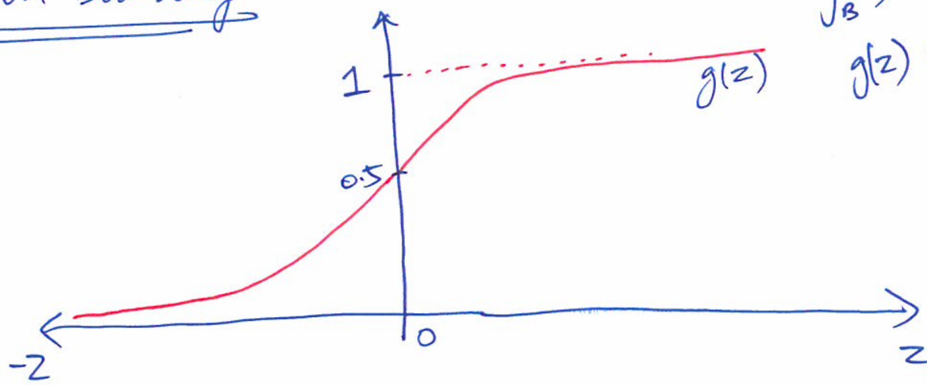
70% of chance that the email (x) is a spam. $\Rightarrow y=1$.

$$P(y=0|B,x) = 1 - P(y=1|B,x).$$

since $y \rightarrow 0$ or 1.

Decision Boundary:

$$\hat{y}_B(x) = g(B^T.x)$$
$$g(z) = \left(\frac{1}{1+e^{-z}}\right)$$



$$g(z) \geq 0.5 \text{ when } z \geq 0$$

$$\boxed{\hat{y}_B(x) = g(B^T.x) \geq 0.5} \text{ when } \boxed{B^T.x \geq 0.}$$

That is,

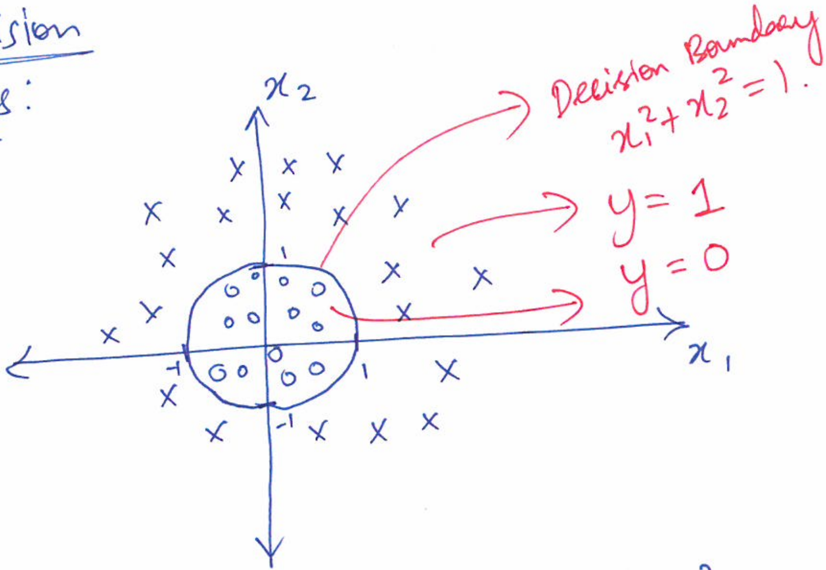
predict "y=1", if $\hat{y}_B(x) \geq 0.5$

$$\boxed{B^T.x \geq 0}$$

predict "y=0", if $\hat{y}_B(x) < 0.5$

$$\boxed{B^T.x < 0}$$

Non-linear Desision Boundaries:



$$\hat{y}_B(x) = g\left(b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2\right).$$

Eg.:- $B = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ Predict "~~y=1~~" if $-1 + x_1^2 + x_2^2 \geq 0$

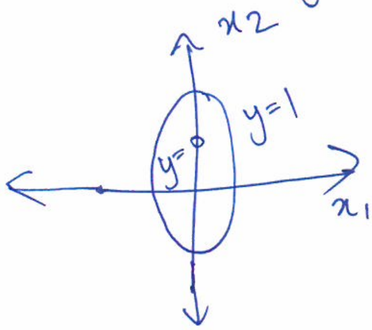
$$\downarrow$$

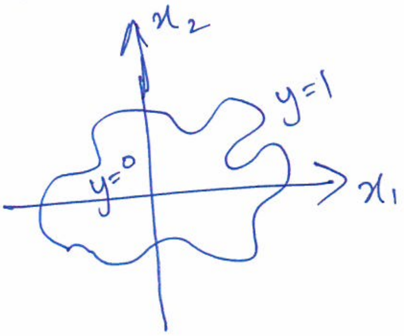
$$\boxed{x_1^2 + x_2^2 \geq 1}$$

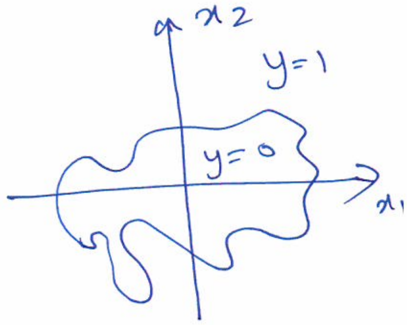
plot $\boxed{x_1^2 + x_2^2 = 1}$ $\rightarrow$ will form a circle.

More complex DB:

$$\hat{y}_B(x) = g\left(b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_1^2 x_2 + \dots\right)$$



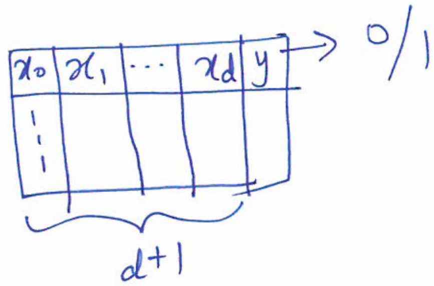




Error fn : (Cost fn.)

$$\hat{y}_B(x) = g(b_0 x_0 + b_1 x_1 + b_2 x_2 + \dots + b_d x_d)$$

$$= g(B^T \cdot x)$$



$$g(z) = \left(\frac{1}{1+e^{-z}}\right)$$

find the values for $B = [b_0, b_1, \dots, b_d]$?

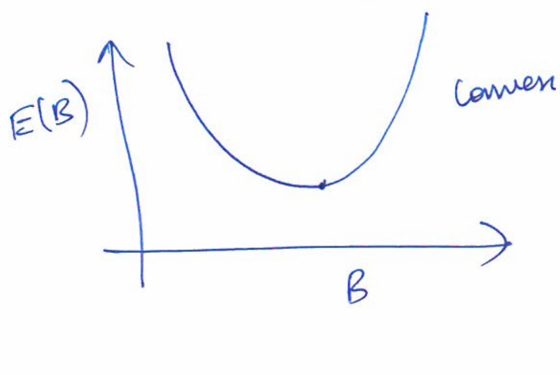
In linear reg, Error fn is,

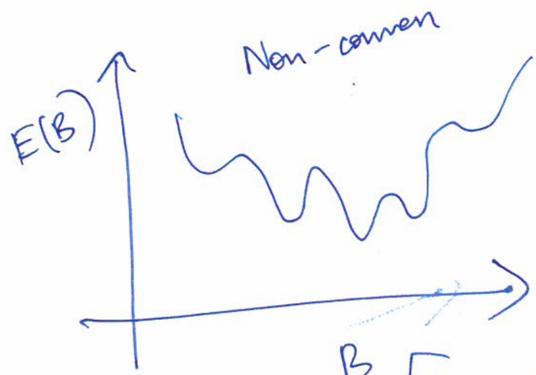
$$E(B) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\frac{1}{2}\left(\overset{B^T \cdot x}{\hat{y}_B^{i}} - y_i\right)^2}_{cost(B,\, (x_i, y_i))}$$

Ok for Linear reg.

In logistic reg,

$$\hat{y}_B(x) = g(z) = \left(\frac{1}{1+e^{-z}}\right)$$





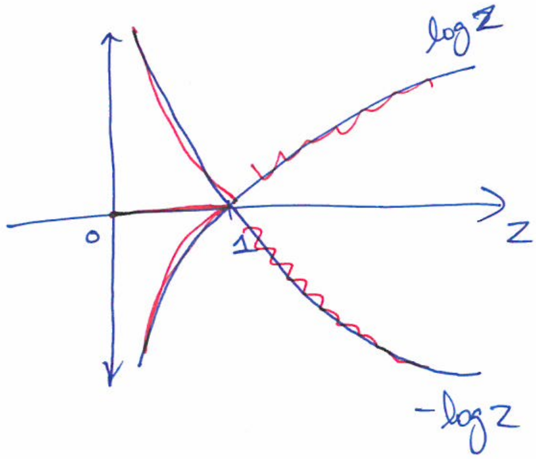
[so we need to find a different cost/Error fn. for logistic reg.]

Cost fn. for Logistic Reg:

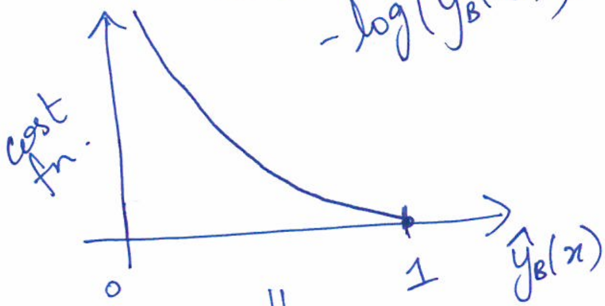
$$\text{cost}\left(\hat{y}_B(x_i),\, y_i\right) = \begin{cases} -\log\left(\hat{y}_B(x_i)\right), & \text{if } y_i = 1, \\ -\log\left(1-\hat{y}_B(x_i)\right), & \text{if } y_i = 0. \end{cases}$$

$\hat{y}_B(x_i)$: predicted class value (compute with B, x_i)

y_i: actual class value (data)



if $\boxed{y_i = 1}$



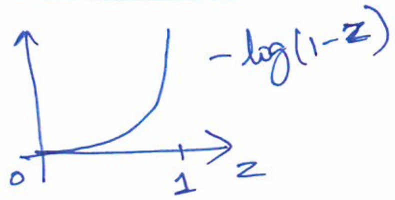
$\Downarrow$

Properties of Cost fn:

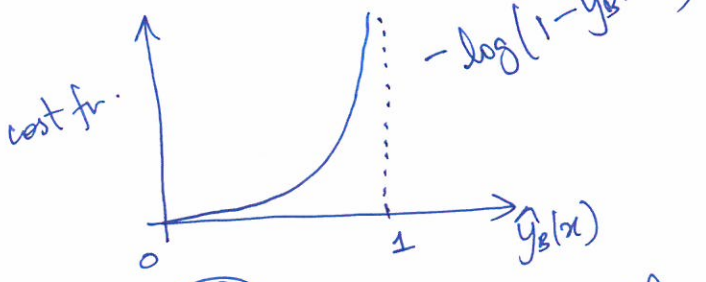
1. Cost $= 0$ if $y_i = 1$ and $\hat{y}_B(x_i) = 1$ } if we correctly predict y, then the cost/error $= 0$.

$P(y=1 \mid B, x_i)$

when $\hat{y}_B(x) \to 0$
cost $\to \infty$ } if we mispredict the y, then cost is ∞.



if $\boxed{y_i = 0}$



2. Cost $= 0$ if $y_i = 0$ and $\hat{y}_B(x_i) = 0$ } if we correctly predict the y, then cost is 0

when $\hat{y}_B(x) \to 1$
cost $\to \infty$ } if we mispredict the y, then cost is ∞.

Rewrite the Cost fn :

$$E(B) = \frac{1}{N} \sum_{i=1}^{N} Cost\left(\hat{y}_B(x_i), y_i\right)$$

for all data samples

$$cost(\hat{y}_B, y_i) = \begin{cases} -\log(\hat{y}_B) & \text{if } y=1 \\ -\log(1-\hat{y}_B) & \text{if } y=0. \end{cases}$$

output variable $y = 0$ or 1.

$$\boxed{cost(\hat{y}_B, y_i) = -y \log(\hat{y}_B(x)) - (1-y) \log(1-\hat{y}_B(x))}$$

$$E(B) = \frac{-1}{N}\left[\sum_{i=1}^{N} y_i \log(\hat{y}_B(x_i)) + (1-y_i)\log(1-\hat{y}_B(x_i))\right]$$

To fit parameter B,

$$\min_B E(B) \longrightarrow \text{Get } B^* ?$$

using GD / SGD.

class value ?

To make a prediction, given a new data x,

Output $\hat{y}_B(x) = g(B^T.x) = \left(\frac{1}{1+e^{-B^T.x}}\right)$

B^T: GD/SGD; x: given.

$$P(y=1 \mid B, x) \Rightarrow \begin{cases} \geq 0.5 \Rightarrow \hat{y}_B(x) = 1 \\ < 0.5 \Rightarrow \hat{y}_B(x) = 0. \end{cases}$$

G. Descent:

$$E(B) = \frac{-1}{N}\left[\sum_{i=1}^{N} y_i \log(\hat{y}_B(x_i)) + (1-y_i)\log(1-\hat{y}_B(x_i))\right]$$

we need find $B^* = \underset{B}{\operatorname{argmin}}\, E(B)$

Repeat until convergence {

$$b_j := b_j - \alpha \frac{\partial E(B)}{\partial b_j}$$ //simultaneous update.

(for $j = 0, 1, 2, \ldots, d$)

}

$$\boxed{\frac{\partial E(B)}{\partial b_j} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_B(x_i) - y_i)\, x_{ij}}$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_d \end{bmatrix}$$

$\frac{\partial E(B)}{\partial b_j} \longrightarrow$ looks same as Lin Reg.

but $\hat{y}_B(x_i) = B^T \cdot x$ in lin. reg.

for logistic reg,

$$\hat{y}_B(x_i) = g(B^T \cdot x) = \left(\frac{1}{1+e^{-B^T \cdot x}}\right)$$

Note: $E(B) \longrightarrow$ error fn. over all training sample (N)

$\text{cost}(\hat{y}_B(x_i), y_i) \longrightarrow$ error fn. for i^{th} training sample.