Binary Classification:

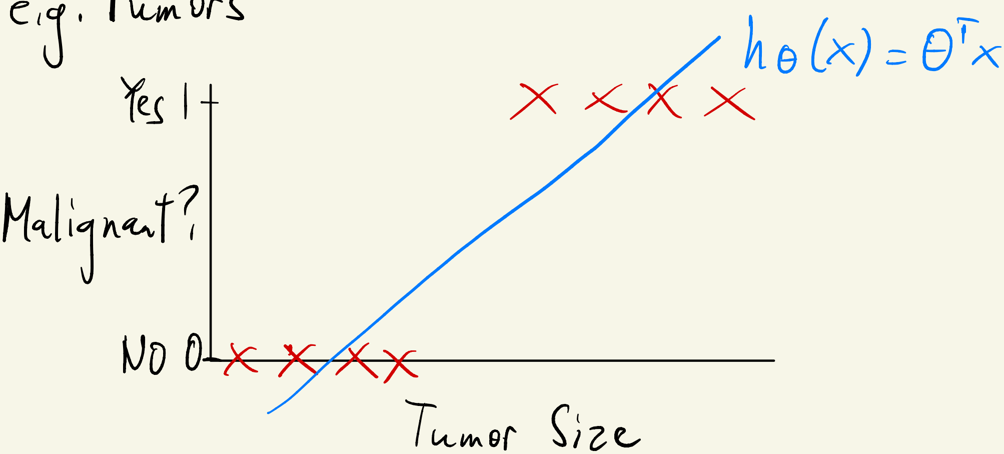
May 14, 2021

In binary classification, the predicted value y can only take on two discrete values: 0 or 1.

i.e. $y \in \{0, 1\}$

→ 0 is the negative class
→ 1 is the positive class

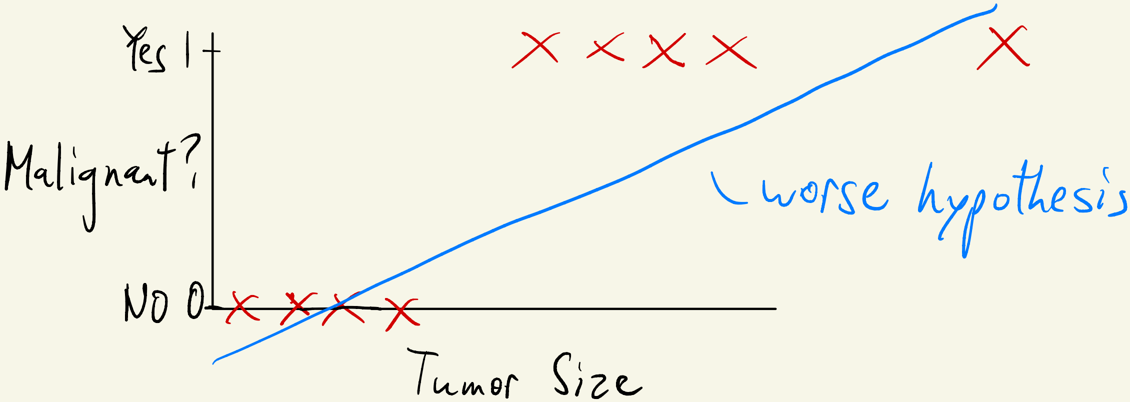
e.g. Tumors



Threshold Classifier Output $h_\theta(x)$ at 0.5:

$$\left.\begin{array}{l} \rightarrow \text{If } h_\theta(x) \geq 0.5,\ y = 1 \\ \rightarrow \text{If } h_\theta(x) < 0.5,\ y = 0 \end{array}\right\} \text{predicted}$$

But data can skew linear regression!



∴ Using linear regression for a classification problem usually isn't a good idea.

✱ Also, if you are using linear regression, the output can be <0 or >1, even if all the training examples only have 0 or 1.

↳ Hence we use the logistic regression algorithm, which makes $h_\theta(x) \in [0, 1]$

i.e. $0 \le h_\theta(x) \le 1$

Logistic Regression Model:

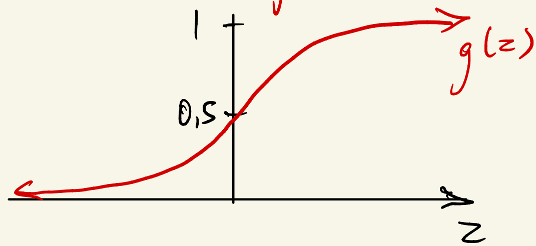
We want $0 \le h_\theta(x) \le 1$.

Let $h_\theta(x) = g(\theta^T x) = g(z)$, where:

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

↳ This is the logistic function, or sigmoid function.

$$\therefore \boxed{h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}}$$



Interpretation of Hypothesis Output:

$h_\theta(x)$ is the estimated probability that $y=1$ on input x.

e.g. $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

$h_\theta(x) = 0.7$, let's say

$\therefore$ Tumor has a 70% change of being malignant.

$\therefore \boxed{h_\theta(x) = P(y=1 \mid x; \theta)}$

↳ the probability that $y=1$ given x parameterized by θ

e.g. If $P(y=1 \mid x; \theta) = 0.7$,
then $P(y=0 \mid x; \theta) = 0.3$.

Since $P(y=1 \mid x; \theta) + P(y=0 \mid x; \theta) = 1$.

Decision Boundary:

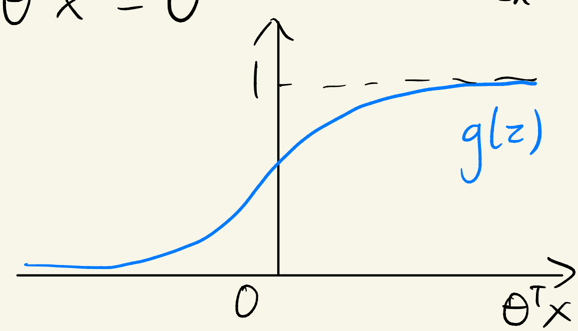
The decision boundary is the line that separates the area where $y=0$ and $y=1$. It is created by $h_\theta(x)$.

e.g.

Predict $y=1$ if $h_\theta(x) \geq 0.5$ and $y=0$ if $h_\theta(x) < 0.5$.

↳ $h_\theta(x) \geq 0.5$ when $\theta^T x \geq 0$

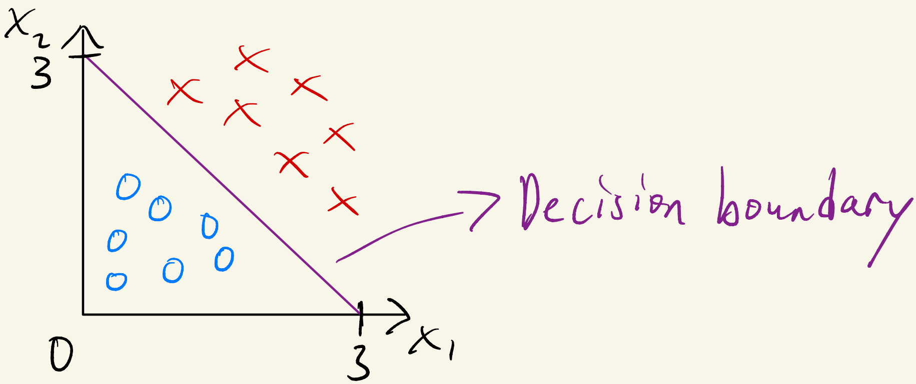
↳ $h_\theta(x) < 0.5$ when $\theta^T x < 0$



e.g. $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

and let $\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \rightarrow \begin{matrix} \theta_0 = -3 \\ \theta_1 = 1 \\ \theta_2 = 1 \end{matrix}$

Predict $y=1$ if $-3 + x_1 + x_2 \geq 0$.



$\therefore$ $x_1 + x_2 = 3$ is the decision boundary

↳ i.e. Like $h_\theta(x) = 0.5$

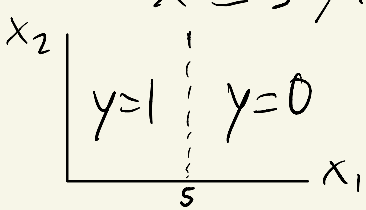
e.g. $\theta_0 = 5$, $\theta_1 = -1$, $\theta_2 = 0$, $h_\theta(x) = g(5 - x_1)$

↳ $y=1$ if $5 - x_1 \geq 0$

$5 \geq x_1$

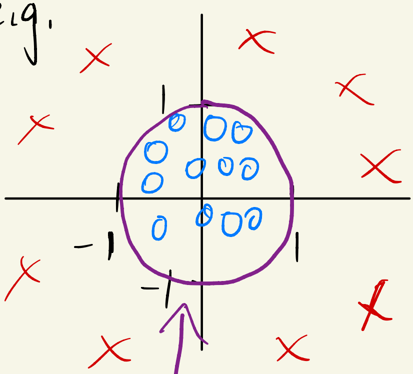
$x \leq 5$, then $y=1$

$\therefore$



Non-Linear Decision Boundaries!

e.g.



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

let $\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

Predict $y=1$ if $-1 + x_1^2 + x_2^2 \geq 0$

↳ $x_1^2 + x_2^2 \geq 1$, circle

You can also have higher-order polynomial decision boundaries.

e.g.