

# Classification and Representation

## Classification

Examples of Classifications problems:

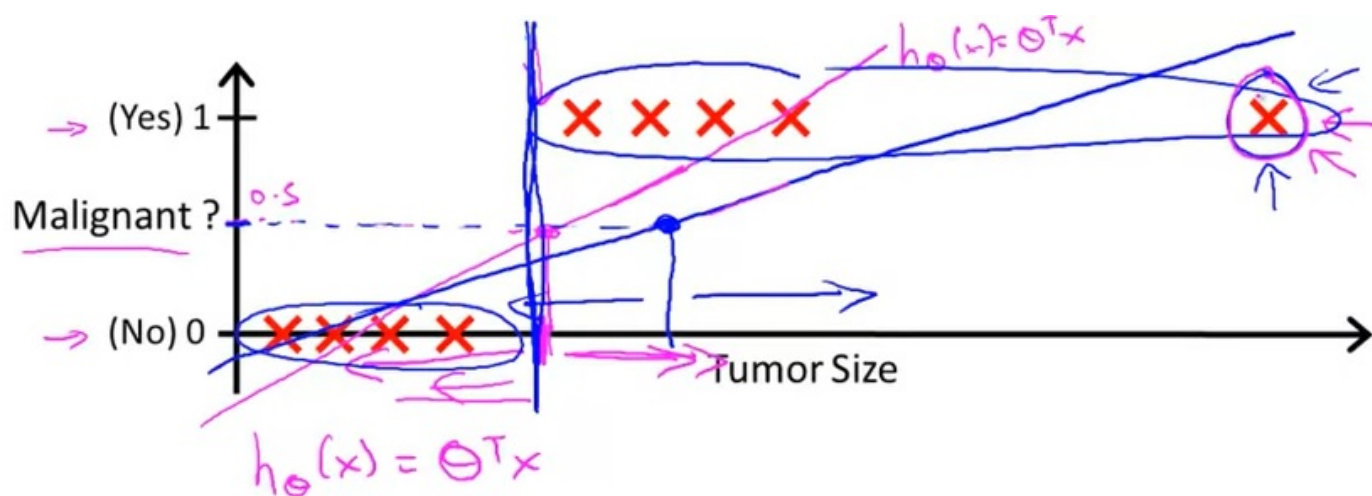
- Email: Spam / Not Spam
- Online Transactions: Fraudulent (Yes/No)
- Tumor: Malignant / Benign

$$y \in \{0, 1\} \quad (0: \text{Negative class}, 1: \text{Positive class})$$

### How do we develop a classification algorithm?

I think that we could do is to apply the algorithm that we already know:

$$h_{\theta}(x) = \theta^T * x$$



If you want to make predictions, one thing that we could do is to use a threshold classifier output at 0.5:

$$h_{\theta}(x) = \begin{cases} 1 & \text{if } h_{\theta}(x) \geq 0.5 \\ 0 & \text{if } h_{\theta}(x) < 0.5 \end{cases}$$

**Note:** For classification we know that  $y$  is either zero or one. But if you are using linear regression where the hypothesis can output values that are much larger than one or less than zero, even if all of your training examples have labels  $y$  equals zero or one.

## Hypotheses Representation

We would like our classifier to output values between zero and one.

When we were using linear regression our hypothesis had the form of:

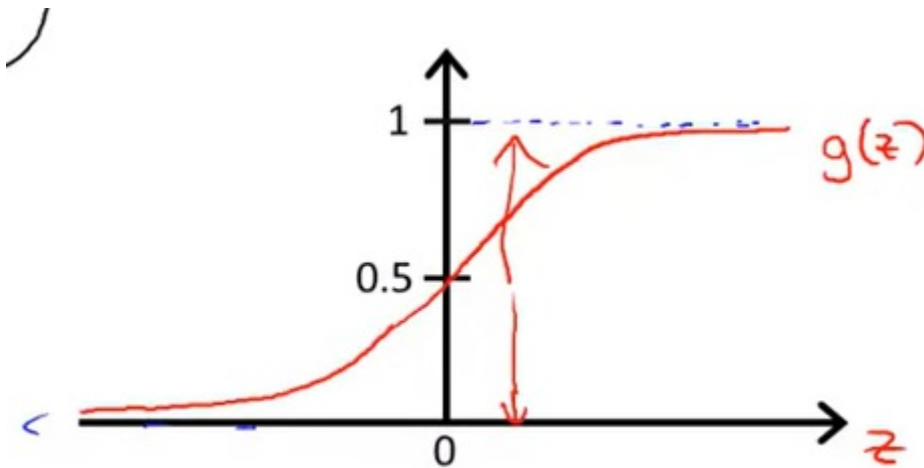
$$h_{\theta}(x) = \theta^T * x$$

For logistic regression we will modify this a little bit and make the hypothesis  $g$  of theta transpose  $x$ .

$$h_{\theta}(x) = g(\theta^T * x), \text{ where } g(z) = \frac{a}{1 + e^{-z}}$$

## Sigmoid \ Logistic Function

The sigmoid function, while it asymptotes at one and asymptotes at zero, as a  $z$  axis, the horizontal axis is  $z$ . As  $z$  goes to minus infinity,  $g(z)$  approaches zero. And as  $g(z)$  approaches infinity,  $g(z)$  approaches one. And so because  $g(z)$  upwards values are between zero and one, we also have that  $h(x)$  must be between zero and one.



## Interpretation of Hypothesis Output

$h_{\theta}(x)$  = estimated probability that  $y=1$  on input  $x$

Example:

$$\text{if } x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ tumorSize \end{bmatrix}$$

$$h_{\theta}(x) = 0.7$$

$h_{\theta}$  will give us the probability that our output is 1. For example,  $h_{\theta} = 0.7$  gives us a probability of 70% that our output is 1. Our probability that our prediction is 0 is just the complement of our probability that it is 1 (e.g. if probability that it is 1 is 70%, then the probability that it is 0 is 30%).

$h_{\theta}(x) = p(y = 1|x; \theta)$  probability that  $y=1$ , given  $x$ , parameterized by  $\theta$

$$h_{\theta}(x) = P(y = 1|x; \theta) = 1 - P(y = 0|x; \theta)$$

$$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$$

## Decision Boundary

In order to get our discrete 0 or 1 classification, we can translate the output of the hypothesis function as follows:

$$h_{\theta}(x) \geq 0.5 \rightarrow y = 1$$

$$h_{\theta}(x) < 0.5 \rightarrow y = 0$$

The way our logistic function  $g$  behaves is that when its input is greater than or equal to zero, its output is greater than or equal to 0.5:

$$g(z) \geq 0.5$$

$$\text{when } z \geq 0$$

Remember.

$$z = 0, e^0 = 1 \Rightarrow g(z) = 1/2$$

$$z \rightarrow \infty, e^{-\infty} \rightarrow 0 \Rightarrow g(z) = 1$$

$$z \rightarrow -\infty, e^{\infty} \rightarrow \infty \Rightarrow g(z) = 0$$

So if our input to  $g$  is  $\theta^T X$ , then that means:

$$h_{\theta}(x) = g(\theta^T x) \geq 0.5$$

$$\text{when } \theta^T x \geq 0$$

From these statements we can now say:

$$\theta^T x \geq 0 \Rightarrow y = 1$$

$$\theta^T x < 0 \Rightarrow y = 0$$

The **decision boundary** is the line that separates the area where  $y = 0$  and where  $y = 1$ . It is created by our hypothesis function.

**Example:**

$$\theta = \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix}$$

$$y = 1 \text{ if } 5 + (-1)x_1 + 0x_2 \geq 0$$

$$5 - x_1 \geq 0$$

$$-x_1 \geq -5$$

$$x_1 \leq 5$$

In this case, our decision boundary is a straight vertical line placed on the graph where  $x_1 = 5$ , and everything to the left of that denotes  $y = 1$ , while everything to the right denotes  $y = 0$ .

Again, the input to the sigmoid function  $g(z)$  (e.g.  $\theta^T X$ ) doesn't need to be linear, and could be a function that describes a circle (e.g.  $z = \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2$ ) or any shape to fit our data.

## Non-linear decision boundaries

