Decision Boundary ¶

What we want to do now is try to understand better when the hypothesis function will make predictions that y is equal to 1 versus when it might make predictions that y is equal to 0. And understand better what hypothesis function looks like particularly when we have more than one feature

$$\rightarrow h_\theta(x) = g(\theta^T x) = P(y=1|x;\theta)$$

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

So what we're shown is that a hypothesis is gonna predict y equals 1 whenever $\theta^T X$ is greater than or equal to 0.

Suppose predict "$y = 1$" if $h_\theta(x) \geq 0.5$

$\rightarrow \theta^T x \geq 0$

$g(z) \geq 0.5$ when $z \geq 0$

$h_\theta(x) = g(\theta^T x) \geq 0.5$ whenever $\theta^T x \geq 0$

z

predict "$y = 0$" if $h_\theta(x) < 0.5$

$h_\theta(x) = g(\theta^T x)$ $\quad g(z) < 0.5$

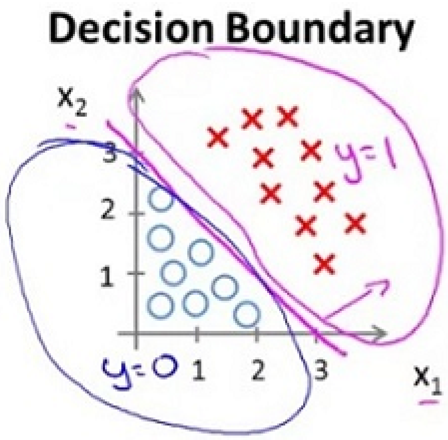
$\theta^T x < 0$

If we decide to predict whether $y = 1$ or $y = 0$ depending on whether the estimated probability is greater than or equal to 0.5, or whether less than 0.5, then that's the same as saying that when we predict $y = 1$ whenever $\theta^T X$ is greater than or equal to 0. And we'll predict y is equal to 0 whenever $\theta^T X$ is less than 0.

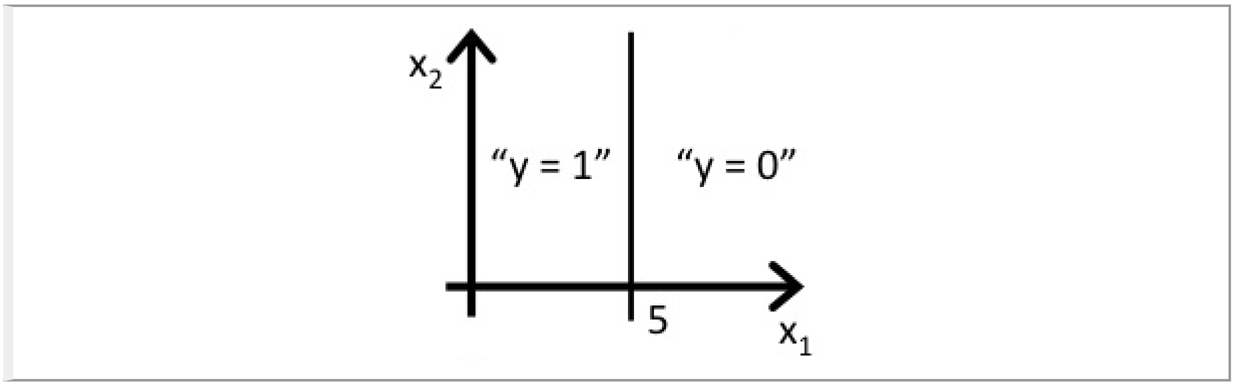
Let's suppose we have a training set like that shown below, and suppose a hypothesis is $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$ and suppose that via a procedure to specified. We end up choosing the following values for the parameter $h_\theta(x) = g((-3) + (1)x_1 + (1)x_2)$. So this means that our parameter vector is going to be $\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$ Predict "$y = 1$" if $-3 + x_1 + x_2 \geq 0$.

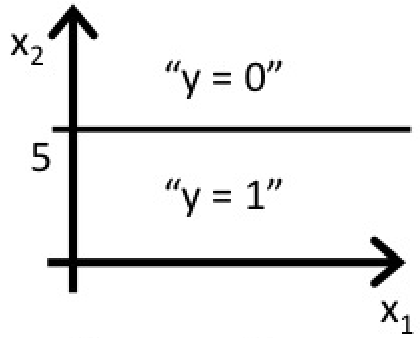
So for any example which features x_1 and x_2 that satisfy $-3 + x_1 + x_2 \geq 0$, our hypothesis will predict that y is equal to 1. We can also take -3 and bring this to the right and rewrite this as $x_1 + x_2 \geq 3$, so equivalently, we found that this hypothesis would predict $y = 1$ whenever $x_1 + x_2$ is greater than or equal to 3.

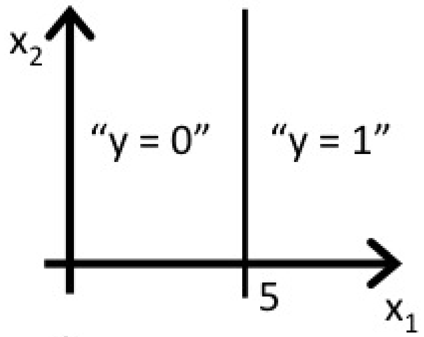
The equation $x_1 + x_2 = 3$ defines the equation of a straight line (The magenta line is called the decision boundary, $x_1 + x_2 = 3 \rightarrow h_\theta(x) = 0.5$).

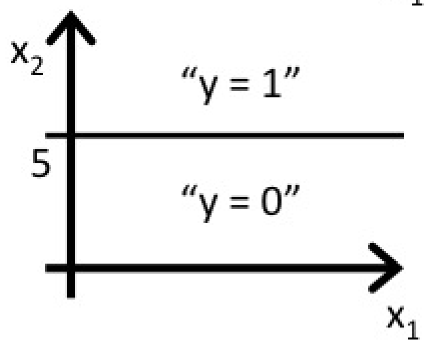


Video Question: Consider logistic regression with two features x_1 and x_2. Suppose $\theta_0 = 5$, $\theta_1 = -1$, $\theta_2 = 0$, so that $h_\theta(x) = g(5 - x_1)$. Which of these shows the decision boundary of $h_\theta(x)$?



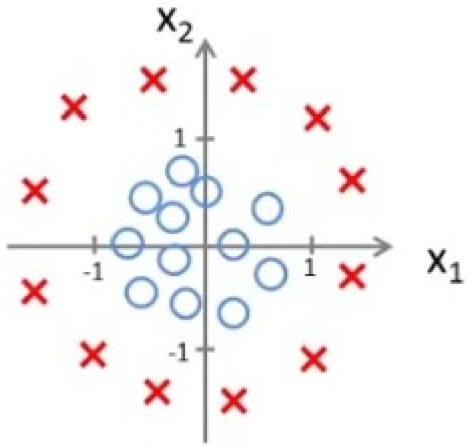
- 

- 

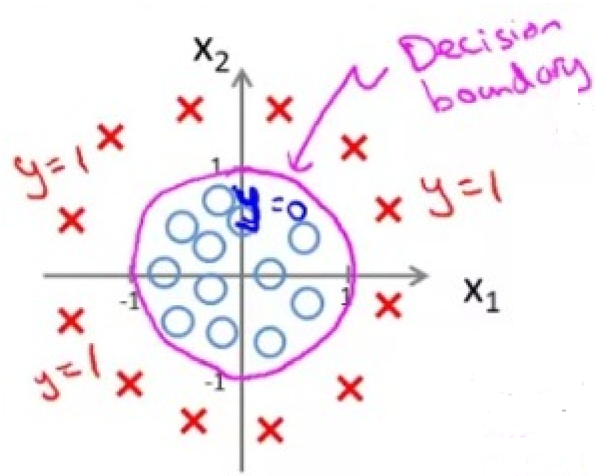
- 

Predict $y = 0$ if x_1 is greater than 5.

Given a training set like the following, how can we get logistic regression to fit the sort of data?, the way to solve this problem is by adding extra higher order polynomial terms to the features. And we can do the same for logistic regression, concretely, let's say our hypothesis looks like:
$h_\theta(x) = g(\theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 x_1^2 + \theta_4 x_2^2)$



suppose that via a procedure to specified. We end up choosing the following values for the parameter theta $h_\theta(x) = (-1) + (0)x_1^2 + (0)x_2^2 + (1)x_1^2 + (1)x_2^2)$. So this means that our parameter vector is going to be $\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Predict "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$ or predict "$y = 1$" if $x_1^2 + x_2^2 \geq -1$.



So by adding more complex, or polynomial terms to our features, we can get more complex decision boundaries that don't just try to separate the positive and negative examples in a straight line, and we can get a decision boundary that's a circle (the decision boundary is a property, not of the trading set, but of the hypothesis under the parameters).

Summary

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:

$g(z) \geq 0.5$

when $z \geq 0$

Remember.

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

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

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

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

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

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:

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

$y = 1$ 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.