Hypothesis Representation ¶

We'd like to come up with a hypothesis that satisfies this property:

Want $0 \leq h_\theta(x) \leq 1$

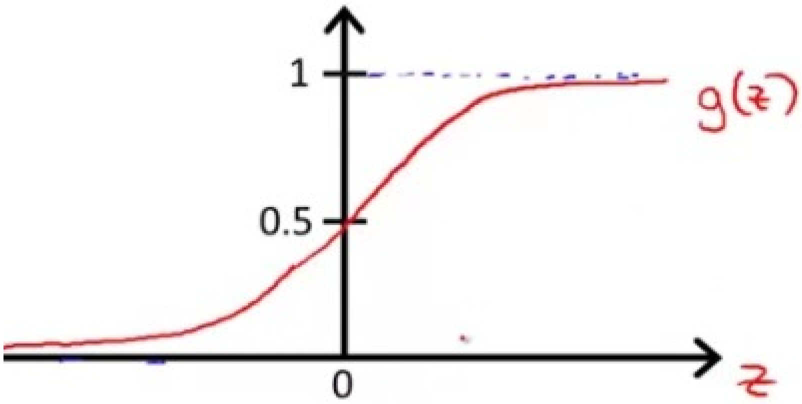
When we were using linear regression, this was the form of a hypothesis, where $h(x)$ is θ transpose x.

$h_\theta(x) = \theta^T x$

For logistic regression, we're going to modify the hypothesis function to make:

$g = (\theta^T x)$

Where we define g like: $g = \frac{1}{1+e^{-z}}$



This is called the sigmoid function, or the logistic function, and the term logistic function, that's what gives rise to the name logistic regression. The terms sigmoid function and logistic function are basically synonyms and mean the same thing. So the two terms are basically interchangeable.

An alternative way of writing out the form of the hypothesis function is:

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

Given the hypothesis function, what we need to do, as before, is fit the parameters theta to our data. So given a training set we need to a pick a value for the parameters theta and the hypothesis function will then let us make predictions.

Interpretation of Hypothesis Output

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

Example: Let's say we're using the tumor classification example, so we may have a feature vector x. And then one feature is the size of the tumor, and we can predict whether a patient's tumor is benign or malignant according to the result obtained by the hypothesis function using the size of the patient's tumor as an input variable.

Example: If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

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

Tell patient that 70% chance of tumor being malignant

We're going to interpret the hypothesis function as:

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

"probability that $y = 1$, given x, parameterized by θ"

and we know that y must be either 0 or 1, where

- $P(y = 0|x; \theta) + P(y = 1|x; \theta) = 0.7 = 1$
- $P(y = 0|x; \theta) = 1 - P(y = 1|x; \theta)$

Video Question: Suppose we want to predict, from data x about a tumor, whether it is malignant ($y = 1$) or benign ($y = 0$). Our logistic regression classifier outputs, for a specific tumor $h_\theta(x) = P(y = 1|x; \theta) = 0.7$ so we estimate that there is a 70% chance of this tumor being malignant. What should be our estimate for $P(y = 0|x; \theta)$, the probability the tumor is benign?

> $P(y = 0|x; \theta) = 0.3$

- $P(y = 0|x; \theta) = 0.7$
- $P(y = 0|x; \theta) = 0.7^2$
- $P(y = 0|x; \theta) = 0.3 \times 0.7$

Summary

We could approach the classification problem ignoring the fact that y is discrete-valued, and use our old linear regression algorithm to try to predict y given x.

However, it is easy to construct examples where this method performs very poorly. Intuitively, it also doesn't make sense for $h_\theta(x)$ to take values larger than 1 or smaller than 0 when we know that $y \in 0, 1$. To fix this, let's change the form for our hypotheses $h_\theta(x)$ to satisfy $0 \leq h_\theta(x) \leq 1$. This is accomplished by plugging $\theta^T x$ into the Logistic Function.

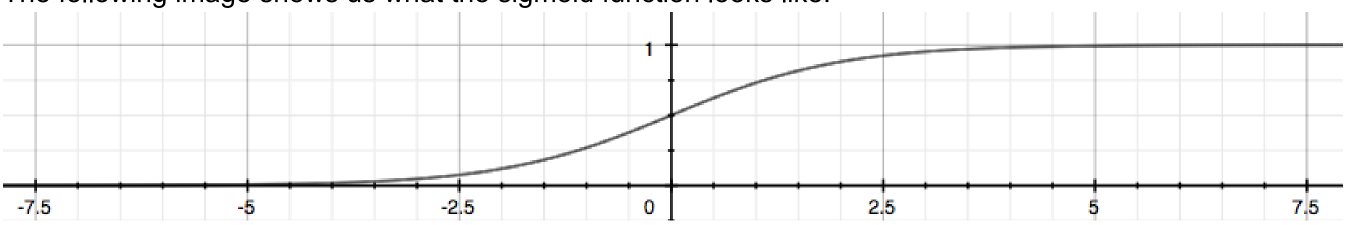
Our new form uses the "Sigmoid Function," also called the "Logistic Function":

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

$x = \theta^T x$

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

The following image shows us what the sigmoid function looks like:



The function $g(z)$, shown here, maps any real number to the $(0, 1)$ interval, making it useful for transforming an arbitrary-valued function into a function better suited for classification. $h_\theta(x)$ will give us the probability that our output is 1. For example, $h_\theta(x) = 0.7$ 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%).

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

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