Hypothesis Representation

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 $h_{ heta}(x)$ 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 $h_{ heta}(x)$ to satisfy $0 \le h_{\theta}(x) \le 1$. This is accomplished by plugging $\theta^T x$ into the Logistic Function.

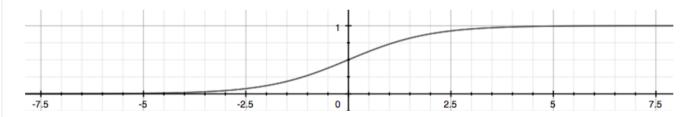
Our new form uses the "Sigmoid Function," also called the "Logistic Function":

$$h_{ heta}(x) = g(heta^T x)$$

$$z = \theta^T a$$

$$z = heta^T x \ g(z) = rac{1}{1 + e^{-z}}$$

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%).

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

✓ Complete

