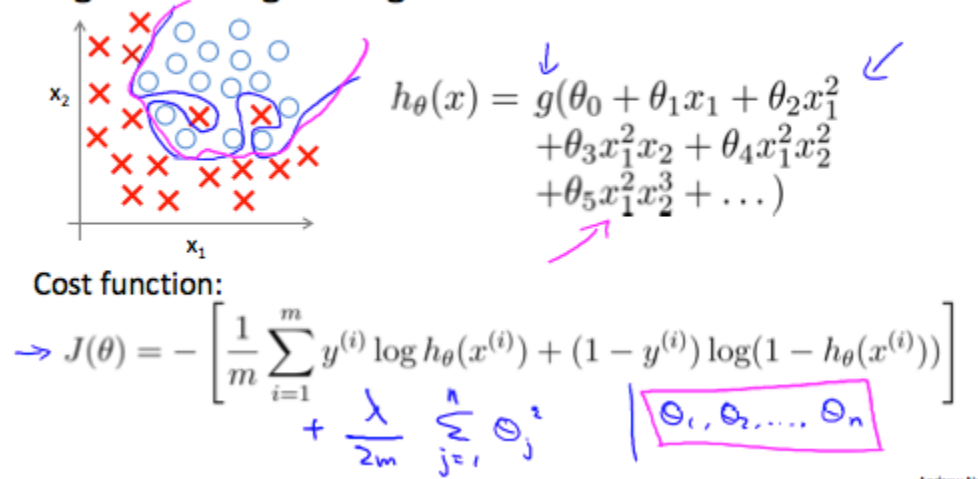# Regularized Logistic Regression

We can regularize logistic regression in a similar way that we regularize linear regression. As a result, we can avoid overfitting. The following image shows how the regularized function, displayed by the pink line, is less likely to overfit than the non-regularized function represented by the blue line:

**Regularized logistic regression.**

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$
$$+ \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2$$
$$+ \theta_5 x_1^2 x_2^3 + \dots)$$

Cost function:

$$J(\theta) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))\right]$$
$$+ \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2 \qquad \boxed{\theta_1, \theta_2, \dots, \theta_n}$$

Andrew Ng

## Cost Function

Recall that our cost function for logistic regression was:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

We can regularize this equation by adding a term to the end:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

The second sum, $\sum_{j=1}^{n} \theta_j^2$ **means to explicitly exclude** the bias term, $\theta_0$. I.e. the θ vector is indexed from 0 to n (holding n+1 values, $\theta_0$ through $\theta_n$, and this sum explicitly skips $\theta_0$, by running from 1 to n, skipping 0. Thus, when computing the equation, we should continuously update the two following equations:

## Gradient descent

Repeat {

$\rightarrow \quad \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$

$\rightarrow \quad \theta_j := \theta_j - \alpha \left[ \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m} \theta_j \right] \leftarrow$

$\qquad\qquad\qquad\qquad (j = \textcolor{red}{\times}, 1, 2, 3, \ldots, n)$

$\qquad\qquad\qquad\qquad\qquad \theta_1 \ldots \theta_n$

}

$\frac{\partial}{\partial \theta_j} J(\theta)$

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