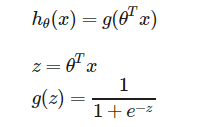
# Logistic Regression

# Classification

The classification problem is just like the regression problem, except that the values we now want to predict take on only a small number of discrete values.

# Hypothesis Representation

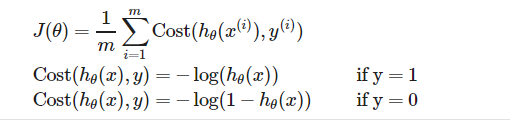
SIGMOD

# Decision Boundary

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

# Cost function

cost function for logistic regression looks like:



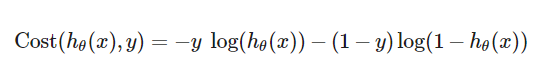
When y = 1, we get the following plot for J(\theta)*J*(*θ*) vs h\_\theta (x)*hθ*​(*x*):



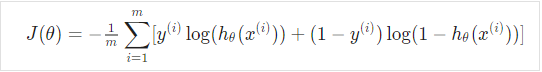
Similarly, when y = 0, we get the following plot for J(\theta)*J*(*θ*) vs h\_\theta (x)*hθ*​(*x*):



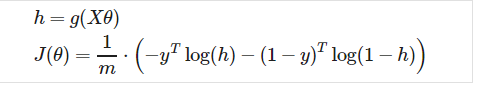
# Simplification：



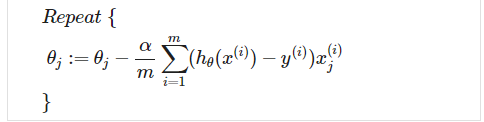
So entire cost function ：



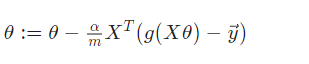
A vectorized implementation is:



将其代入梯度下降公式去：



A vectorized implementation is:



# 其他几个advanced optimization

"Conjugate gradient", "BFGS", and "L-BFGS" are more sophisticated, faster ways to optimize θ that can be used instead of gradient descent.

学习Octave调用方法。

# Multiclass Classification: One-vs-all

# cqmPjanSEeawbAp5ByfpEg_299fcfbd527b6b5a7440825628339c54_Screenshot-2016-11-13-10.52.29

# Overfittingg0cOOdKsMEeaCrQqTpeD5ng_2a806eb8d988461f716f4799915ab779_Screenshot-2016-11-15-00.23.30

we’ll say the figure on the left shows an instance of underfitting—in which the data clearly shows structure not captured by the model—and the figure on the right is an example of overfitting.

This terminology is applied to both linear and logistic regression. There are two main options to address the issue of overfitting:

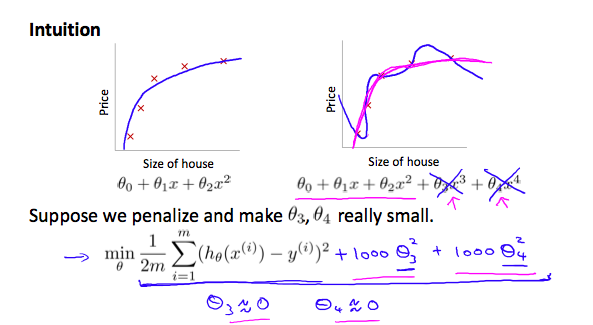
1) Reduce the number of features:

* Manually select which features to keep.
* Use a model selection algorithm (studied later in the course).

2) Regularization

* Keep all the features, but reduce the magnitude of parameters *θj*​.
* Regularization works well when we have a lot of slightly useful features.

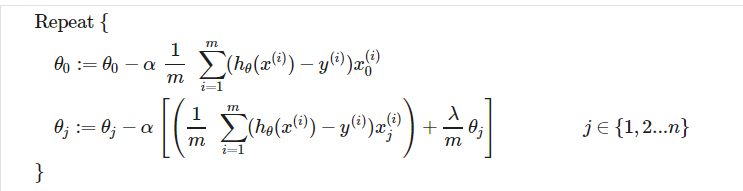
# Cost function

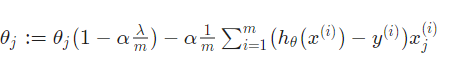


We could also regularize all of our theta parameters in a single summation as:

1571314745(1)

The λ, or lambda, is the ****regularization parameter****. It determines how much the costs of our theta parameters are inflated.





To add in regularization, the equation is the same as our original, except that we add another term inside the parentheses:

