机器学习 week3

目录

1 Classification **2**

2 Hypothesis Representation **3**

1 Classification

To attempt classification, one method is to use linear regression and map all predictions greater than 0.5 as a 1 and all less than 0.5 as a 0. However, this method doesn't work well because classification is not actually a linear function.

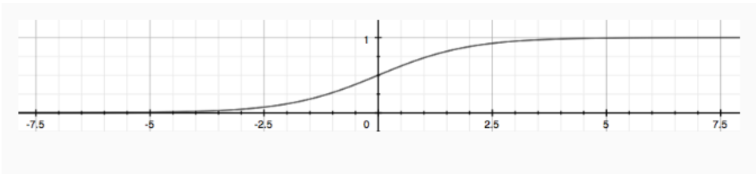
The classification problem is just like the regression problem, except that the values we now want to predict take on only a small number of discrete values. For now, we will focus on the **binary classification** problem in which y can take on only two values, 0 and 1. (Most of what we say here will also generalize to the multiple-class case.) For instance, if we are trying to build a spam classifier for email, then [Math Processing Error] may be some features of a piece of email, and y may be 1 if it is a piece of spam mail, and 0 otherwise. Hence, y 0,1. 0 is also called the negative class, and 1 the positive class, and they are sometimes also denoted by the symbols "-" and "+." Given [Math Processing Error], the corresponding [Math Processing Error] is also called the label for the training example.

2 Hypothesis Representation

We could approach the classification problem ignoring the fact that y is discrete-valued,and use our old linear regression algotithm 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[Math Processing Error]to take values larger than 1 or smaller than 0 when we know that y 0, 1.To fix this,let's change the form for our hypothese[Math Processing Error]to satisfy[Math Processing Error].This is accomplished by plugging [Math Processing Error]into the Logistic Function.

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

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.

[Math Processing Error]will give us the **probability** that our output is 1.For example,[Math Processing Error]gives us a probability of 70% that our outputs 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 30%).