Note: Logistic Regression

Sun Zhao

December 24, 2012

1 Classification

In classification problem, we will predict the label for the given data instead of continuous values like regression problem. For example, a classifier may automatically label an email as spam or not spam, an online transaction fraudulent or not fraudulent, or a tumor malignant or benign. In the simplest case, there are only two classes. Usually, we use 0 and 1 to denote the negative class and the positive class separately. Figure1 shows a training data set of tumor classification. When using linear regression algorithm and exclude blue cross example, we get a hypothesis of purple line. Naturally, we define 0.5 as the boundary of malignant or benign, ie, if the hypothesis h(x) is equal to or greater than 0.5, then x is malignant, otherwise benign. Let $h(x_0) = 0.5$, then if x ¿= x_0, x is malignant, otherwise benign. In this case, we get a perfect classifier. However, if the blue cross is included, we may get a hypothesis of green line. And let $h(x_1) = 0.5$, obviously, x_1 is worth than x_0 when classifying tumors. This is because in the binary classification problem, the value of y is always 0 or 1, but the output of linear regression can be larger than 1 or smaller than 0. Logistic regression will use a hypothesis $h_\theta(x)$ whose output is always between 0 and 1.

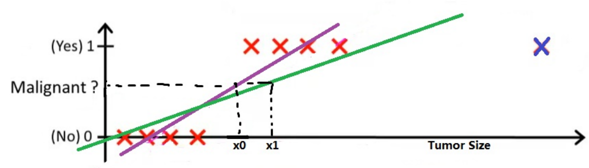

Figure 1:

2 Hypothesis Representation

How to control the output of our hypothesis between 0 and 1? The answer is using a sigmoid function shown in formula1. The sigmoid function plotted in Figure2 maps an interval of (-∞, ∞) to (0, 1). Combining the sigmoid function and hypothesis function $h_\theta(x) = \theta^T x$, we will get the hypothesis function of logistic regression shown in formula2. The output of the new hypothesis $h_\theta(x)$ means the probability that y = 1 on input of x. For example, in the tumor classification problem, if $h_\theta(x_0) = 0.7$, then we can tell the patient whose tumor size is equal to x_0 that there are 70% chance that her/his tumor to be malignant. Formally, the hypothesis predicts the probability of x being positive class given x and θ. This definition is described in formula3, and obviously, we can induce a property shown in formula4.

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{2}$$

$$h_\theta(x) = P(y = 1|x; \theta) \tag{3}$$

$$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1 \tag{4}$$

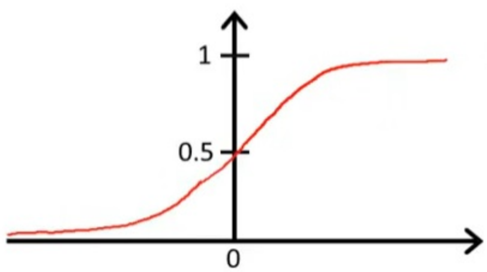

Figure 2: