Note: Linear Regression Model

Sun Zhao

December 15, 2012

1 Model Representation

Remind that regression algorithms are tying to infer a function predicting continuous values when given training samples. Let's first focus on the notation of the training samples. Regarding to Table1, training samples are represented by a vector of $\mathbf{x}$ and a vector of $\mathbf{y}$ with identical length of m and each sample becomes a pair of ($x^{(i)}$, $y^{(i)}$). The inferred function h(which means hypothesis) is shown as below:

$$h_\Theta(x) = \Theta_0 + \Theta_1 x \tag{1}$$

For simplicity, formula(1) consists of only one variable, thus we call this linear regression as univariate linear regression.

2 Cost Function

Θ_0 and Θ_1 are the two arguments of hypothesis function, so the idea here is to choose optimal values of Θ_0 and Θ_1. Intuitively, the optimal h should minimize average the error distance between the predict values of each $x^(i)$ and $y^(i)$. Then we can define the cost function as below:

$$J(\Theta_0, \Theta_1) = \frac{1}{2m}\sum_1^m (h_\Theta(x^{(i)}) - y^{(i)})^2 \tag{2}$$

Obviously, the object function is defined as:

$$\underset{\Theta_0,\Theta_1}{minimize}\frac{1}{2m}\sum_1^m (h_\Theta(x^{(i)}) - y^{(i)})^2 \tag{3}$$

Using the cost function, we can easily calculate the cost value of any hypothesis functions. For example, Figure1 shows three samples(cross pointers) and a hypothesis function $h(x) = 0.5x$. The cost value of it is calculated as:

$$J(0.5) = \frac{1}{2 \times 3}[(0.5 - 1)^2 + (1 - 2)^2 + (1.5 - 3)^2] = \frac{3.5}{6} \tag{4}$$

3 Gradient Descent

Now we turn to the calculation methods of the objection function. Suppose we have a training set as shown in Figure3. We can enumerate all the possible values of Θ_0 and Θ_1 and calculate the corresponding cost $J(\Theta_0, \Theta_1)$ and show them in a 3D space like Figure2. In Figure2, if we start from a random pointer($\Theta_0, \Theta_1, J(\Theta_0, \Theta_1)$), we can iteratively migrate a small step through the negative direction of the gradient direction of this point and hopefully, we will converge to the lowest point of the surface.