Regularization

Zongheng Yang

May 19, 2012

1 Overfitting and Solutions

Overfitting: fits training data well, yet fails to generalize and has poor performance on test data.

Solutions:

1. Reduce # of features (manually select features; use model selection algorithms).

2. Regularization: reduce maginitudes of all θ, works well when we have a lot of features. If θ's become small, the cost function becomes "smoother".

2 Regularized Linear Regression

In order to minimize the magnitudes of all θ, the cost function is added with a new term:

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

The gradient descent:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}; \tag{2}$$

$$\theta_j := \theta_j - \alpha[\frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} - \frac{\lambda}{m}\theta_j] \tag{3}$$

$$= \theta_j(1 - \alpha\frac{\lambda}{m}) - \alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}. \tag{4}$$

($(1 - \alpha\frac{\lambda}{m})$ is usually sightly smaller than 1, so essentially in regularization θ_j is adjusted to a smaller value.)

The normal equation: if $\lambda > 0$,

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \right)^{-1} X^T y. \tag{5}$$

Notice that if $m(\#examples) \leq n(\#features)$, the original $X^T X$ is non-invertible. However, regularization here guarantees invertibility.

3 Regularized Logistic Regression

Cost function:

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2. \tag{6}$$

The gradient descent has the same form as that of regularized linear regression.

4 θ_0 and λ

θ_0 is usually not regularized.

If λ is too large $\rightarrow$ all θ's are approximately zeros except $\theta_0 \rightarrow h_\theta(x) = \theta_0$; this is called **underfitting**.