Normal Equation

Gordon Gao at 2016.09.08

Instead of using Gradient Descent algorithm, we also can just use a normal equation to find the optimal value for theta all at once. However, Normal Equation has some disadvantages. We will discuss when we can use it, and when we cannot use it.

Normal Equation Function:

$$\theta = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot y$$

X is the features of the training examples, y is the prediction vector of the training examples.

Octave implementation:

$$pinv(X' \cdot X) \cdot X^T \cdot y$$

Advantages:

- No need to choose α.
- Don't need to iterate.

Disadvantages:

- Need to compute $\left(X^T \cdot X\right)^{-1}$, time complexity: $O(n^3)$.
- Slow if n(features) is very large which should use Gradient Descent. if n < 10,000, we can use Normal Equation. Otherwise, we should use Gradient Descent.

What if $\left(X^T \cdot X\right)^{-1}$ is non-invertible?

First, we should use pinv instead of inv in Octave. pinv will deal with the non-invertible matrix.

Second, the model may have too many features(m <= n). Then, we should delete some features or use regularization.