Note: Regularization

Sun Zhao

December 28, 2012

1 Over-fitting

Over-fitting problem occurs when the learned hypothesis fits the training data set very well, but fails to generalize to new examples. It is mainly coursed by the trained model is excessively complex, such as having too many parameters relative to the number of observations. Fig. 1 shows three linear regression hypothesises for the house prices predicting problem. The hypothesis shown in Fig. 1A is called "under-fitting" which means it is quite bias from the right one. Fig. 1C is an example of over-fitting which has high variance. Though it predicts the prices perfectly for every examples in training data, it can not be generalized to new input. Moreover, Fig. 2 shows "under-fitting", "right one", "over-fitting" logistic regression hypothesis separately.
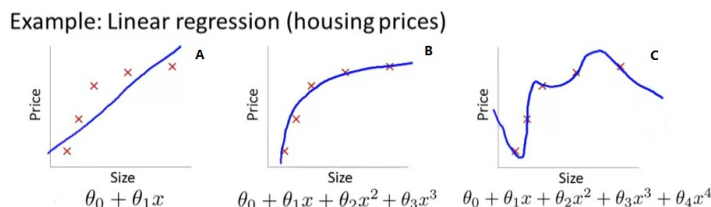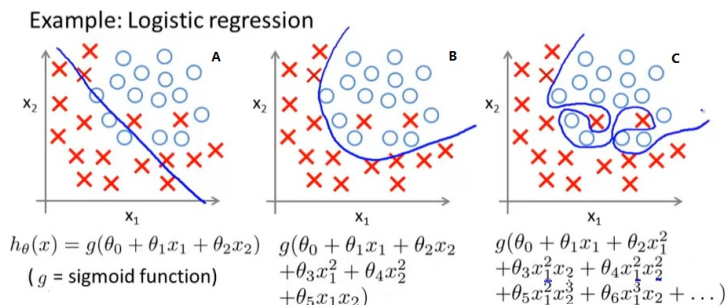

Figure 1:

Figure 2:

2 Regularized Cost Function

The intuition of regularization is to penalize large parameters of θ and keep the hypothesis simple. In Fig. 1C, if we can penalize θ_4 and shrink it to zero, the hypothesis will much closer to the "right" one. Regularization adds a term of

$\frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j$ to the regression cost function. So, the new cost function is as (1).

$$J^{'}(\theta) = J(\theta) + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j \tag{1}$$

The larger value of θ_j, the larger cost of the corresponding hypothesis, hence the gradient descent algorithm will penalize large θ_j.

3 Regularized Gradient Descent