

Regularisation notes.

Robert K.

December 27, 2016

Abstract

This is just a document with notes regarding Regularisation in finding Θ values to prevent *overfitting*.

1 Underfit vs Overfit

Underfit - (a.k.a "high bias") is the prediction (learned hypothesis function) which does not fit the training or new input data very well. Too simplified hypothesis function.

Overfit - (a.k.a "high variance") is the prediction (learned hypothesis function) which fits well the training data, but fails to generalise to new input data. It can occur when we have too many *features* and small size of training set.

How to address overfitting problem:

1. Reduce the number of features, which can be done in two ways:
 - Manually select which features to keep
 - Use Model Selection Algorithm to automatically select features to keep
2. Use **regularisation** where we keep all features, but with different priority

2 Regularisation

Regularisation - is the technique of reducing overfitting by keeping all features but reducing the values of θ_j .

It works well when we have a lot of features and each of them contributes a bit to predicting y .

Properties of regularisation:

- if θ_j is small then more likely we will not overfit. We want to keep θ_j small.

To use Regularisation we slightly modify cost function $J(\theta)$

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

If the θ_j is small, then the cost function $J(\theta)$ will be small too.

λ - is the **regularisation parameter** It determines how much the costs of our theta parameters are inflated.

Properties of new cost function:

- cost will be small if θ_j is small.
- cost will be small for the big **regularisation parameter** λ if θ_j will be super small
- If λ is chosen to be too large, then $\theta_1 \dots \theta_n$ will be near zero to make cost small. It may cause underfitting.
- if $\lambda = 0$ or is too small then it is similar to case of not using regularisation at all - so we increase a chance of overfitting.