

EE655000 Machine Learning

Homework 2 Report

110062802 呂宸漢

1. Maximum Likelihood and Least Squares (MLLS)

這次作業的 spec 中有說明要如何計算 feature vector (Φ)，因此我們只需要將 spec 中計算 Φ 的公式打進程式，再依照課本上的 $w_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t$ (3.15)，將 training data 計算出來的 feature vector 帶入 Φ 並將 training data 的 label 帶入 t 以算出 model 的 weight。在 predict 時，只需要將 testing data 計算出的 Φ 依照 spec 所述 $y(x, w) = \sum_{j=1}^{P+2} w_j \phi_j(x)$ 把 ϕ 與對應的 weight 相乘並加總，即可得到 predict 的機率。

由於 spec 上 feature vector 的擺法為 $\Phi(x) = [\phi_1(x), \phi_2(x), \dots, \phi_{P+2}(x)]^T$ 與課本上的 Φ 差一個 transpose，為了可以直接利用課本的計算公式，我在計算 Φ 時是依照 spec 的順序並按照課本上的擺法擺放。由於 Φ 的維度為 $(n, P+2)$ ，weight 的維度為 $(P+2, 1)$ ，因此在 predict 時可以將計算方式簡化成 $\Phi \times W$ 的矩陣乘法，利用矩陣乘法的特性，便可以直接得到 predict 的結果 y ，維度為 $(n, 1)$ 。

O1 = 2, O2 = 4:

MSE of BLR = 0.0074112157477848484, MSE of MLR = 0.007886485349383129.

2. Bayesian Linear Regression (BLR)

Bayesian linear regression 的 weight 計算方式為 $m_N = S_N(S_0^{-1}m_0 + \beta \Phi^T t)$ ， $S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi$ ，將 $S_0^{-1} = \alpha I$ 帶入化簡後可得 $w = (\lambda I + \Phi^T \Phi)^{-1} \Phi^T t$ (3.28)，

其中 I 為單位矩陣、 $\lambda = \frac{\alpha}{\beta}$ 為常數，透過此公式即可計算 model 的 weight。在了解 weight 的計算方式後便可以利用與 MLLS 相同的流程計算 weight 與 predict testing data 的結果。

O1 = 2, O2 = 4:

MSE of BLR = 0.0074112157477848484, MSE of MLR = 0.007886485349383129.

3. Discuss the difference between MLLS and BLR

由兩者計算 weight 的公式可以發現，其實兩個 function 大同小異，只是 BLR 在計算 weight 的公式多了 λI 這一項，這一項主要是為了不要讓 model 的 weight 過大，避免 model overfitting。當 $\lambda = 0$ 時，BLR 就會與 MLLS 相同，因此可以將 BLR 視為有 regularization 的 MLLS。

The impact of different choices of O1 and O2

由下面兩張圖可以看出無論是在 MLLS 或是在 BLR 的 prediction result 中，當 O1 與 O2 越大時，training 的 MSE 會變得越來越小，testing 的 MSE 則會變大；反之在 O1 與 O2 變小時，training 的 MSE 會變大，testing 的 MSE 則會變小。會造成這種結果是因為在我們的 model 裡面， $O1 \times O2$ 決定了 model 的複雜度，model 越複雜越可以 fit training data，不過也越容易造成 overfitting 的情況。比較 Figure 1 與 Figure 2 的折線可以看出因為 MLLS 本身不會限制 weight 的大小，所以在不同的 O1 與 O2 的設定上 training MSE 都差不多，可是在 O1 與 O2 都較大時 testing MSE 會急速上升，代表 model 已經 overfitting 了；而 BLR 則是因為有 λI 這一項做 regularization，所以雖然 O1 與 O2 變大 testing MSE 有變大，不過就沒有像 MLLS 這麼誇張，有成功控制 weight 的大小，避免 overfitting。在我的 case 中 $O1=2$ 且 $O2=4$ 在兩者有較小的 testing MSE，因此我將 default 的 O1 與 O2 分別設成 2 與 4。

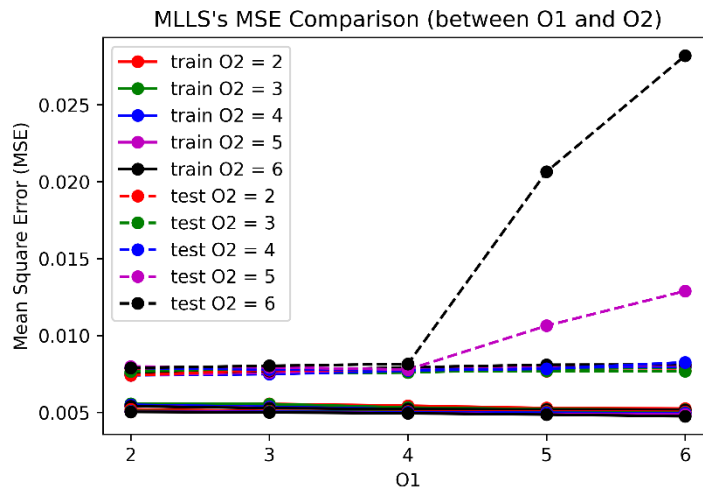


Figure 1: MLLS's MSE comparison

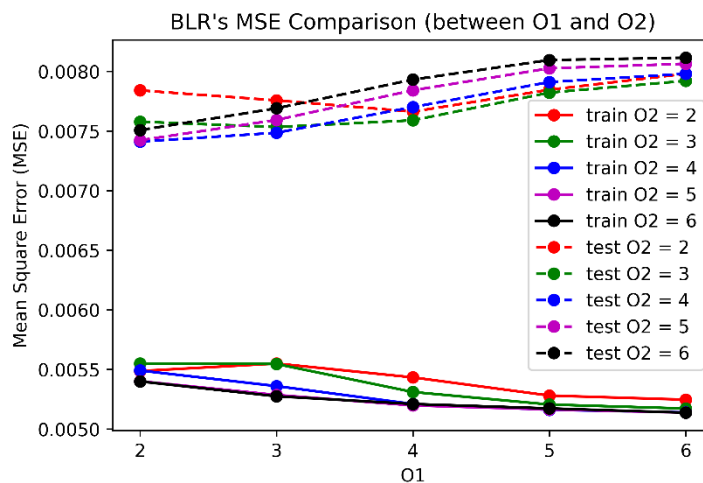


Figure 2: BLR's MSE comparison