

Problem (3)

- Please discuss the difference between Maximum Likelihood and Bayesian Linear Regression.

- Maximum Likelihood 求得 Weight 的方法是透過 least square solution:

$$W = (\Phi(X)^T \Phi(X))^{-1} \Phi(X)^T Y$$

其中 $\Phi(X)$ 求得的方式是透過作業要求的方法(如下圖)

$$\phi_k(\mathbf{x}) = \exp \left\{ -\frac{(x_1 - \mu_i)^2}{2s_1^2} - \frac{(x_2 - \mu_j)^2}{2s_2^2} \right\}, \quad \text{for } 1 \leq i \leq O_1, 1 \leq j \leq O_2,$$

其矩陣大小為 $(n, (O_1 * O_2 + 2))$ ， n 為 training data 數量，而 Y 的矩陣大小為 $(n, 1)$ 。故最後求得的 W 大小為 $((O_1 * O_2 + 2), 1)$ ，預測的方式是先將 testing data 轉換成 $\Phi(X)$ 的形式，並將剛剛訓練好的 W 與其做矩陣相乘 $(\Phi(X) * W)$ ，即可求得該 model 所要的 predict value(size 為 $(n, 1)$)。

- Bayesian Linear Regression 求得 Weight 的方法也類似：

$$W = (\lambda I + \Phi(X)^T \Phi(X))^{-1} \Phi(X)^T Y$$

就是多了 λI 這個 term(其他部分都一樣)， λ 為常數， I 為單位矩陣，這個 term 的主要意義為不希望讓 model weight 的值過大，用以避免 model 對 training data 過擬和(overfitting)，當 λ 值越大，代表 weight value penalty 越大，所求得的 weight 就會越小。故 Bayesian Linear Regression 可以當成是帶有 Regularization term 的 Maximum Likelihood Estimation(MLE)。

- The impact of different choices of O_1 and O_2 and results in your report.

- $O_1=O_2=2$

```
Testing data set : MSE of BLR = 0.007351319037571889, MSE of MLR= 0.007409343830683199.  
Training data set : MSE of BLR = 0.005208990058767792, MSE of MLR= 0.005195351995391876.
```

- $O_1=O_2=5$

```
Testing data set : MSE of BLR = 0.00773107875869716, MSE of MLR= 0.010633785014418114.  
Training data set : MSE of BLR = 0.004979861895070076, MSE of MLR= 0.0048718036180833086.
```

- $O_1=O_2=10$

```
Testing data set : MSE of BLR = 0.01043620427045089, MSE of MLR= 41.77270324717341.  
Training data set : MSE of BLR = 0.00451514787357394, MSE of MLR= 0.0035955745368886614.
```

O_1 & O_2 的大小影響了 $\Phi(X)$ feature size，當 O_1 & O_2 越大，feature size 越大，訓練出來的 weight 也比較多，換句話說就是 model complexity 變高，而過高的 model complexity 會造成 overfitting，也就是在 training data 上的 MSE 雖然很小，但是在 testing data 上的 MSE 可能很大。從上面我做的實驗可發現，當 O_1 & O_2 越大，BLR 與 MLR 對 Training data 的 MSE 都越來越

小，但是 BLR 與 MLR 對 Testing data 的 MSE 卻越來越大，甚至在 $O1=O2=10$ 的時候 MSE 飆很高。明顯驗證我上述所說 **overfitting** 的情況。