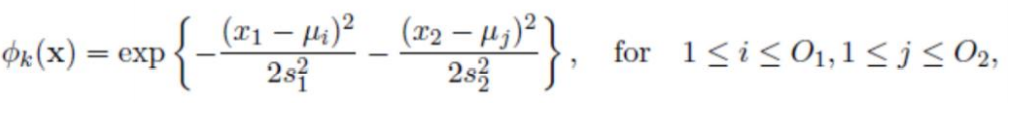
Problem (3)

* 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:

其中求得的方式是透過作業要求的方法(如下圖)

其矩陣大小為(n, (O1\*O2+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((O1\*O2+2), 1)，預測的方式是先將testing data轉換成的形式，並將剛剛訓練好的W與其做矩陣相乘()，即可求得該model所要的predict value(size為(n,))。

* + Bayesian Linear Regression求得Weight的方法也類似：

就是多了這個term(其他部分都一樣)，為常數，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 O1 and O2 and results in your report.
  + O1=O2=2



* + O1=O2=5



* + O1=O2=10



O1 & O2的大小影響了 feature size，當O1 & O2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O1 & O2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。