Maximum Likelihood and Bayesian Linear Regression

從公式來看Maximum Likelihood在預測時他只會去找尋一個最優解的回歸式，但不會考慮到模型的整體複雜度，而Bayesian Linear Regression比Maximum Likelihood多考慮了prior，在假設prior是高斯分布的狀況下可推得在樣本數增加後預測的變數減少，產生一個正則化的作用，而從程式的實驗解果來看，如果將正則化的λ設定為1並且O1和O2都設定為2去做最優化可發現loss值反而比Maximum Likelihood還低，但如果將λ設定為1e-5即可得到比Maximum Likelihood還優的解，而當O1和O2設定為10即可發現Bayesian Linear Regression的loss比Maximum Likelihood小超過一萬倍，因此推測在O1和O2較小的狀況整個模型並不複雜因此有無正則化的結果並不明顯，但假設O1和O2大的話，整體模型複雜的狀況有正則化的優點就十分明顯了。

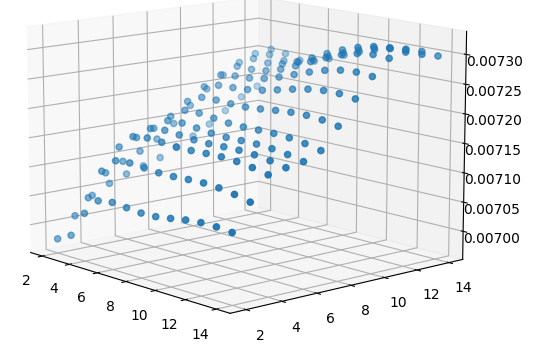
解聯立方程式對比梯度下降

在我們列出公式後我們可以選擇使用解聯立方程式或是梯度下降法來求解最優的參數，梯度下降法可參考hw2\_SGD.py，解聯立方程式可參考hw2.py，在O1和O2同為2的狀況下我們可以發現解聯立方程式可將loss降到0.006但梯度下降只能到0.01，而梯度下降法中模型的參數以隨機均勻分布和batch size設定為1來講結果會最好，不過當O1和O2同為5的時候會發現解聯立方程式的Maximum Likelihood的loss到0.4而Bayesian Linear Regression為0.007，但梯度下降法的結果均為0.01，因此推測梯度下降法要到最優解的狀況比較難，但比較不容易陷入區域最佳解的狀況。

共同高斯分布

在一開始取Feature Vector時需要使用的x的min和max，如果將測試資料的Feature Vector裡面的x的min和max取的跟訓練時一樣可再將loss降低。如果取的是測試資料的min和max在Maximum Likelihood可以得到0.008而如果取一樣可以降到0.006。

O1和O2參數設置

我嘗試窮舉O1和O2 在2~15所有組合，在解聯立方程式的方法下Maximum Likelihood的模型下可觀察到在某些特殊的位置有可能會導致loss暴增，而隨著O1\*O2越大Loss值也越大，最小可以到達0.006，最大可到達7481，而在Bayesian Linear Regression也會有隨著O1\*O2越大loss值也越大的趨勢，但不會發生突然的暴增，平均loss落在0.007，可參考以下圖，xy軸分別代表O1和O2，y代表loss值。因此選擇最優的O1和O2為2。而使用梯度下降法也可觀測到類似遞增數據。