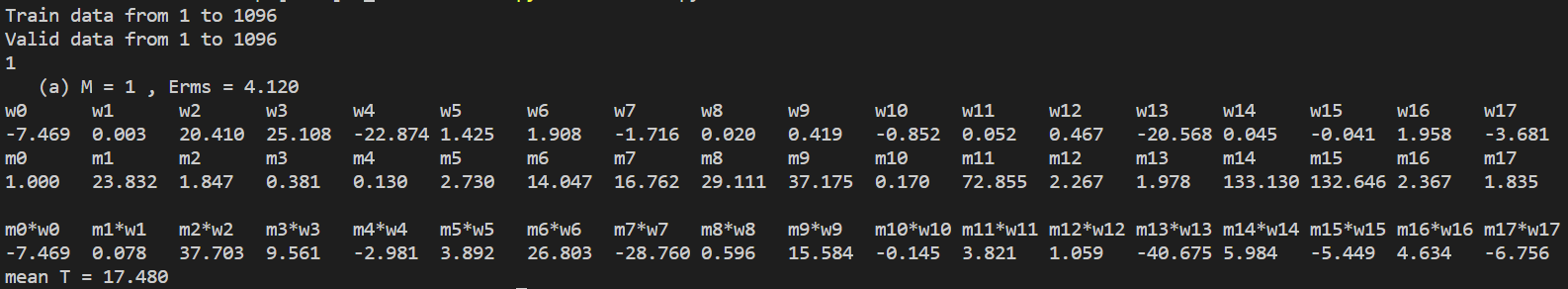
1.Feature selection

(a)

M = 1

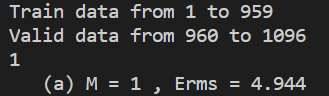
這邊我對於M=1的實作方法為將E(w)直接進行微分求極值。推倒過程如下圖(一)：

1. All input data is training data / Curve fitting (Erms = 4.120)

這部分將所有的data\_X都作為training data，接著再算全部資料的誤差值。

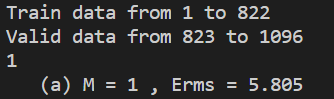
1. 將第960～1096筆資料作為valid data(Erms = 4.944)

縮減training data量，可以發現誤差值提升。



1. 將第823～1096筆資料作為valid data (Erms = 5.805)

繼續縮減training data數量，誤差值明顯提升。



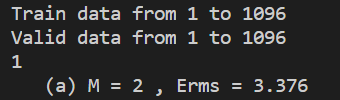
由2～3可以確定，當給定的訓練資料量增大時，RMS的準度會增加是無庸置疑的，只是當資料量一減少，誤差值的差距在M=1的模型會有顯著的差異。

M = 2

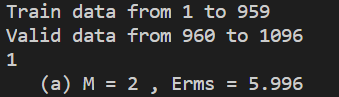
這裡因為考慮到更高維度的參數眾多(171個)，直接對E(w)作微分求極值的方法較為困難。我雖然有先試過直接微分的方式來試著做看看，但參數過多的關係，光是要建立聯立方程式Ax=b的A矩陣就要計算許久，所以最後選擇直接使用老師上課所提到的方法：

直接找出矩陣來做運算，對於程式整體的運算效率與速度會更加的優秀，因此決定改用此方法來實作。而M = 2的部分可以看到若是作為curve fitting時RMS明顯變小，但是當training data逐漸變小時，M = 2的RMS變化會比M = 1的還要劇烈，可以推測應該是M = 2有overfitting的現象發生。

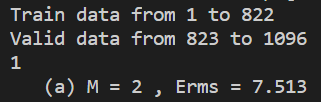
1. All input data is training data / Curve fitting (Erms = 3.376)

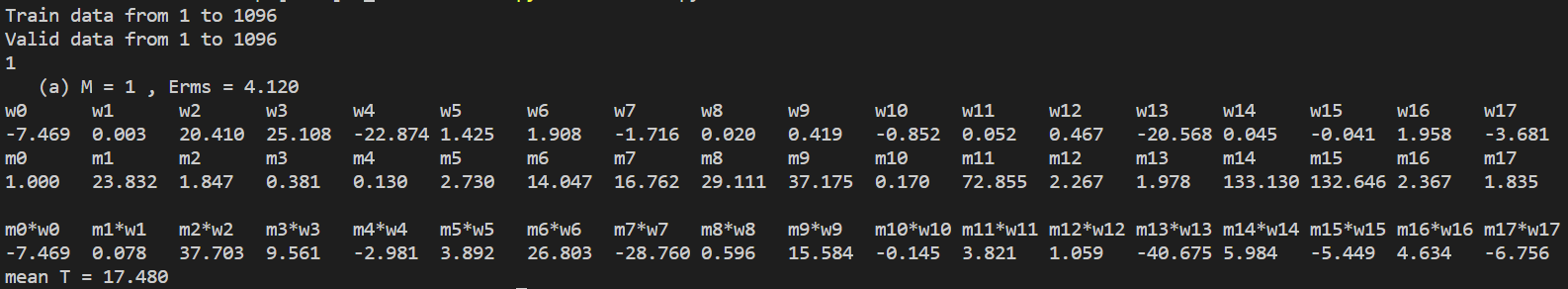


1. 將第960～1096筆資料作為valid data (Erms = 5.996)



1. 將第823～1096筆資料作為valid data (Erms = 7.513)



(b)

上圖w0~w17為參數w向量值，m0~m17為17筆資料的平均值。原先我要進行分析時，只考慮到w的部份，所以會認為w越大表示此項的重要性越高。但當我將此想法套用到第二題，根據我認為重要的w帶入後會發現計算出的rms效果不如我想像中的好。所以我又回過頭來看，認為所預測出的結果是weight與base function線性組合而成，應該要考量兩者相乘後的數值來做判斷。因此，我將各項資料(x1~x17)的平均值與weight相乘，根據這項來作為判斷標準。根據此標準，我認為有以下幾組可能：

1. w2、w3、w4、w5、w6、w7、w9、w11、w12、w13、w14、w15、w16、w17(abs>1)
2. w2、w3、w4、w5、w6、w7、w9、w11、w13、w14、w15、w16、w17(abs>2)
3. w2、w3、w6、w7、w9、w13、w14、w15、w17(abs>5)
4. w2、w6、w7、w9、w13(abs>10)
5. w2、w3、w6、w9、w14(>5)

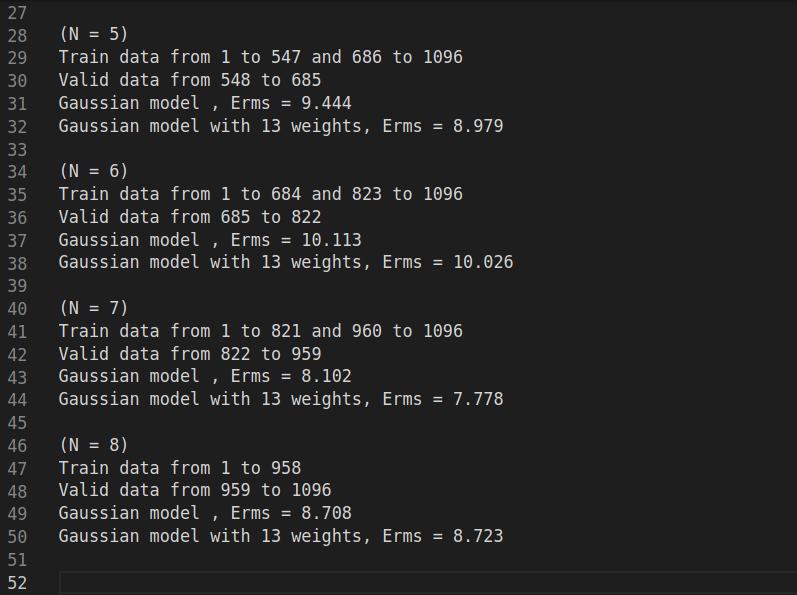
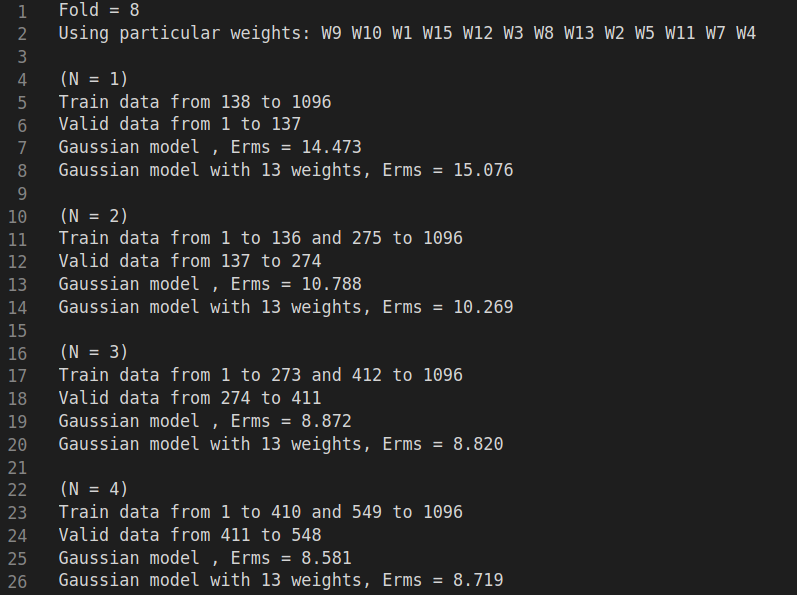
單就直覺而言，我認為選擇絕對值數值越大的加權項對於整體model會越有貢獻，因為當數值越大表示此項只要一有變動便很容易去影響到模型。只是要如何去抉擇threshold的部份，我尚且沒有什麼概念，所以我考慮將每項的數值來取絕對值，再根據不同的threshold來做分類。另外，考量到或許也有可能此模型只要考慮到正向數值即可，所以另外做一類數值大於5的類別。因此，我總共考量5種可能。

但是我將此想法落實到第二題來作為我選擇的依據時，會發現並非如此…...詳細的討論我將寫在第二題。

2.Maximum likelihood approach

此題我使用的基底函數為高斯分佈模型，根據老師上課所提到的分析，可以得知：

根據上面的方程式，可以快速的計算出Gaussian maximum likelihood 的平均值與標準差還有基底函數。首先，我考量所有的資料來做分析，可以得到如下圖的結果(hw2-1\_Output\_.txt)：

我將data\_X分為8份，來進行train與 valid。其中的particular weights 是我將全部data都作為training data 時，一次取一項資料去除來計算rms，計算每項去除後的rms變化，若變化越劇烈的，就將其列入particular weights，我這邊只取到4.作為threshold。

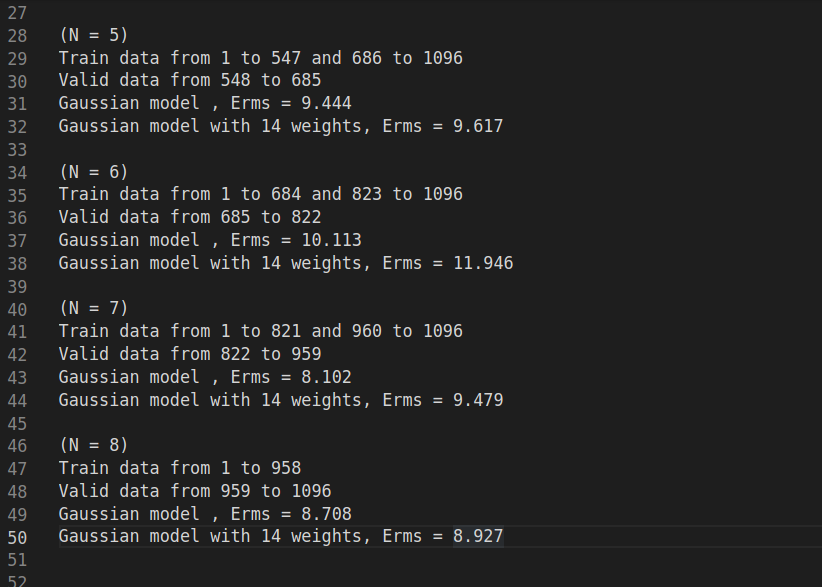
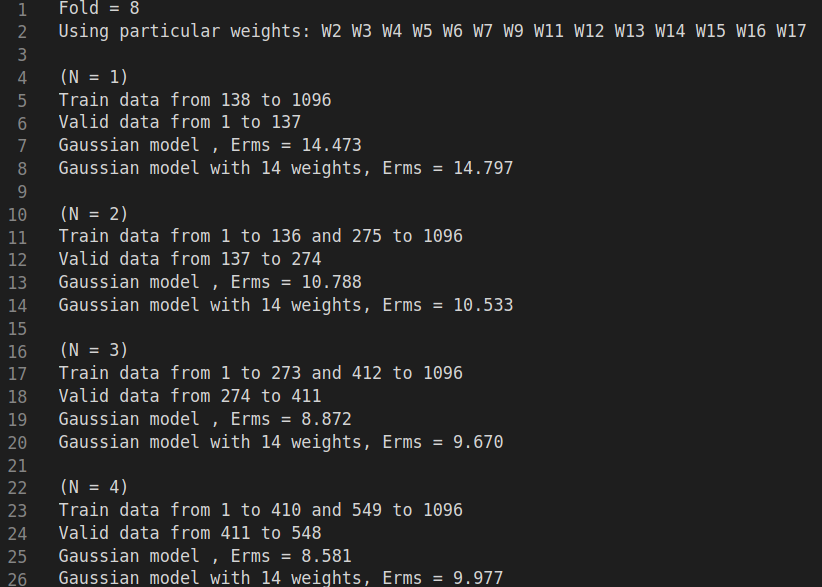
其中各項數值可參考下表：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| original | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

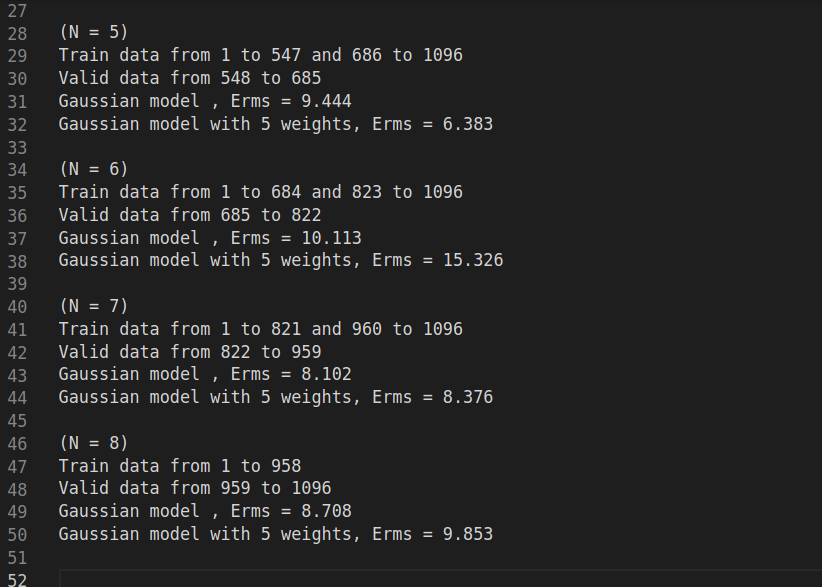
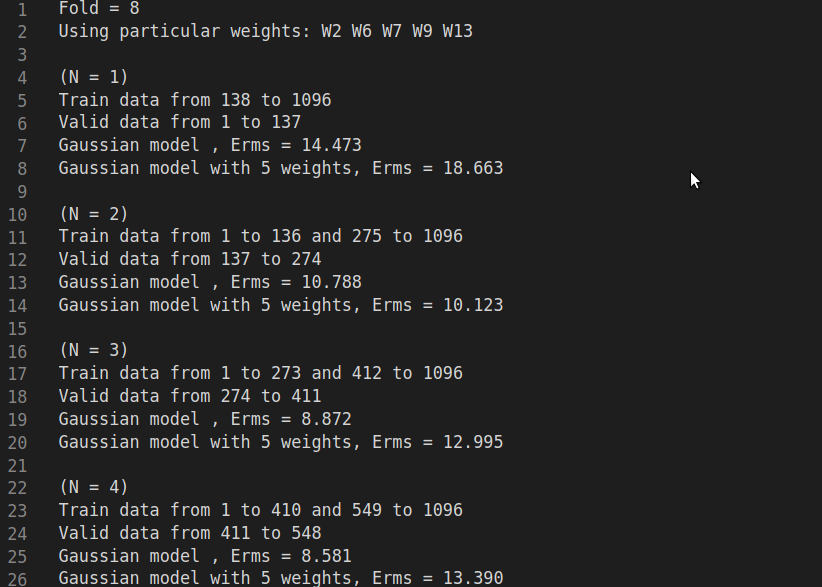
由此可發現，除了N=1,4,8外，只取特定weights來做預測的話，rms的數值會較全部資料都拿來做預測的還要好。在此可以推測全部都資料都拿來作為高斯分佈模型的話，會有overfitting的可能性。

接著，回頭來看在第一題所推估的幾類來作為particular weights的話，rms變化狀況：

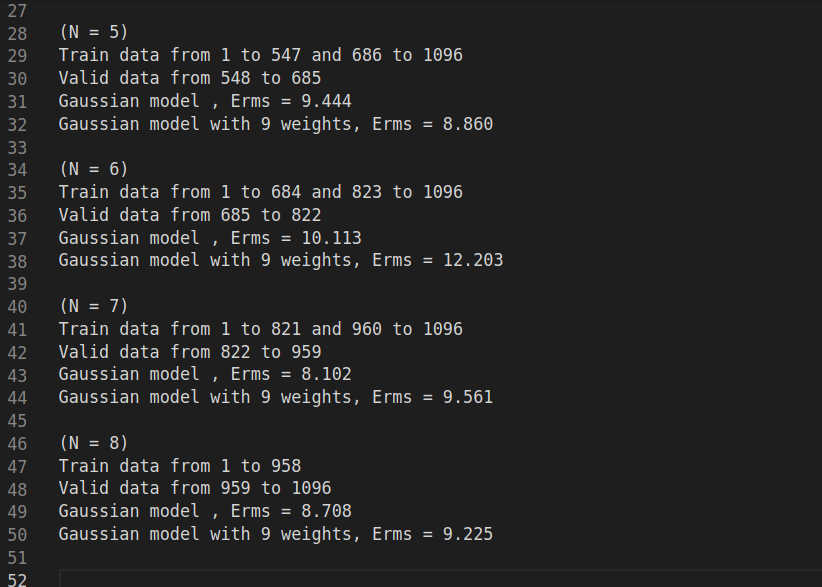
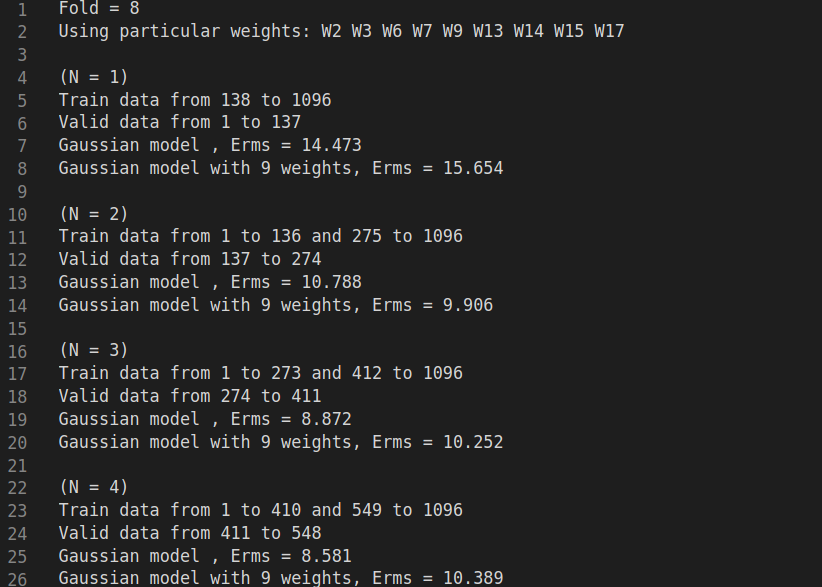
1) w2、w3、w4、w5、w6、w7、w9、w11、w12、w13、w14、w15、w16、w17(abs>1)



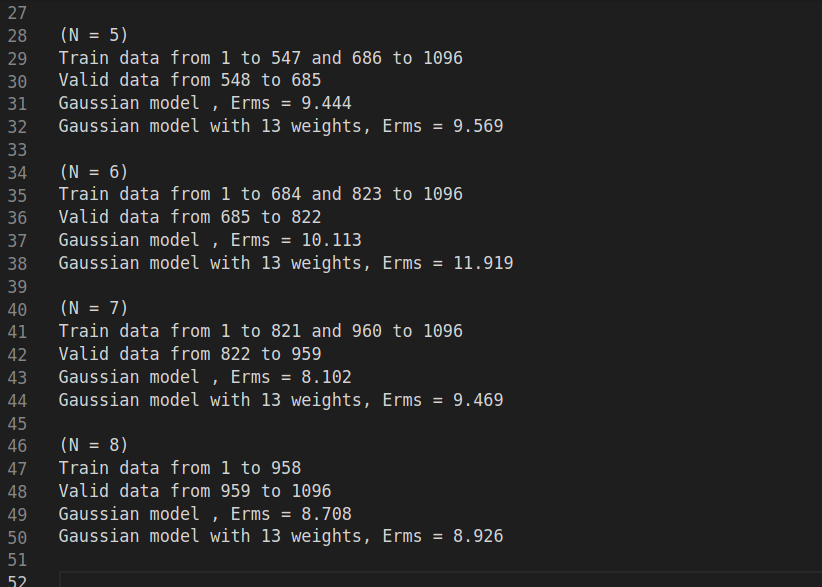
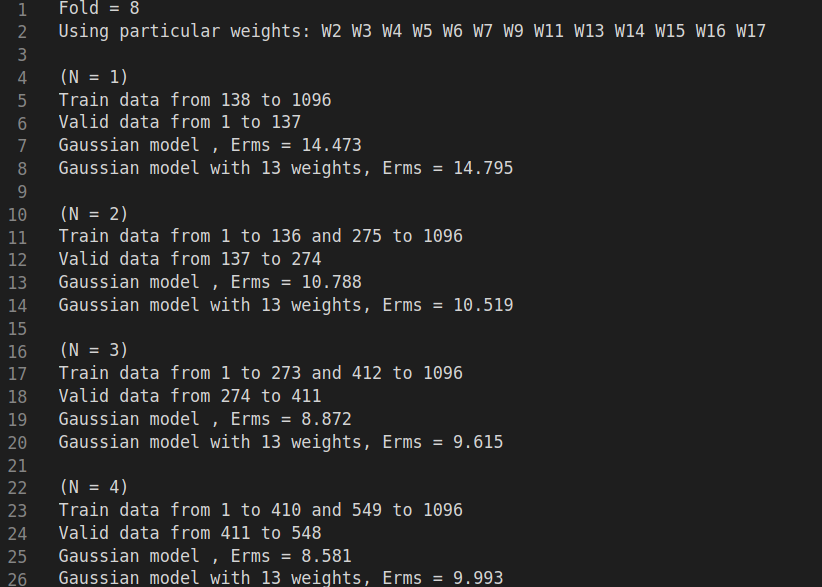
2) w2、w6、w7、w9、w13(abs>10)



3) w2、w3、w6、w7、w9、w13、w14、w15、w17(abs>5)



4) w2、w3、w4、w5、w6、w7、w9、w11、w13、w14、w15、w16、w17(abs>2)



5) w2、w3、w6、w9、w14(>5)

