0881505 曾翊銘

HW 1

1. Bayesian Linear Regression

$p(w|x,t) \propto p(t|x,w)p(w)$

● 證明

● 補充

根據課本

$p(x) = N(x|u, \Lambda^{-1})$

$p(y|x) = N(y|Ax+b, L^{-1})$

$p(y) = N(y|Au+b, L^{-1}+A\Lambda^{-1}A^T)$

$p(x|y) = N(x|\Sigma\{A^T L(y-b)+\Lambda u\}, \Sigma)$, $\Sigma=(\Lambda+A^T L A)^{-1}$

$$\Rightarrow \begin{cases} p(w) = N(w|0, \alpha^{-1}I) & (u=0, \Lambda^{-1}=\alpha^{-1}I) \\ p(t|x,w) = N(t|y(x,w), \beta^{-1}) & (A=\phi^T(x), b=0, L^{-1}=\beta^{-1}) \end{cases}$$

$= w^T\phi(x) = \phi^T(x)w$

結合這兩項

$\Rightarrow p(w|x,t) = N(w|\Sigma\{A^T L(t-b)+\Lambda u\}, \Sigma)$ $S=(\alpha I+\phi(x)\beta\phi^T(x))^{-1}$

$= N(w|S\{\phi(x)\beta t\}, S)$ $=(\alpha I+\beta\sum_{n=1}^{N}\phi(x_n)\phi(x)^T)^{-1}$

結合上面幾項 $\begin{cases} p(w|x,t) = N(w|S\{\phi(x)\beta t\}, S) \\ p(t|x,w) = N(t|\phi^T(x)w, \beta^{-1}) \end{cases}$ $u=S\{\phi(x)\beta t\}$, $\Lambda^{-1}=S$, $A=\phi^T(x), b=0, L^{-1}=\beta^{-1}$

$\Rightarrow p(t|x,\cancel{x},t) = N(t|Au+b, L^{-1}+A\Lambda^{-1}A^T)$

$\Rightarrow N(t|\phi^T(x)S\{\phi(x)\beta t\}+0, \beta^{-1}+\phi^T(x)S\phi(x))$

$\Rightarrow N(t|\beta\phi^T(x)S\sum_{n=1}^{N}\phi(x_n)t_n, \beta^{-1}+\phi^T(x)S\phi(x))$

$\Rightarrow N(t|m(x), s^2(x))$ #

2. Linear Regression

1. Feature selection

(a)小題

模型設置：

	Model 1	Model 2
Basic Function	Polynomial (M = 1)	Polynomial (M = 2)

模型共有設置：

使用特徵維度(D)	17
訓練資料佔比	80%
Error Function	RMSE

實做結果：

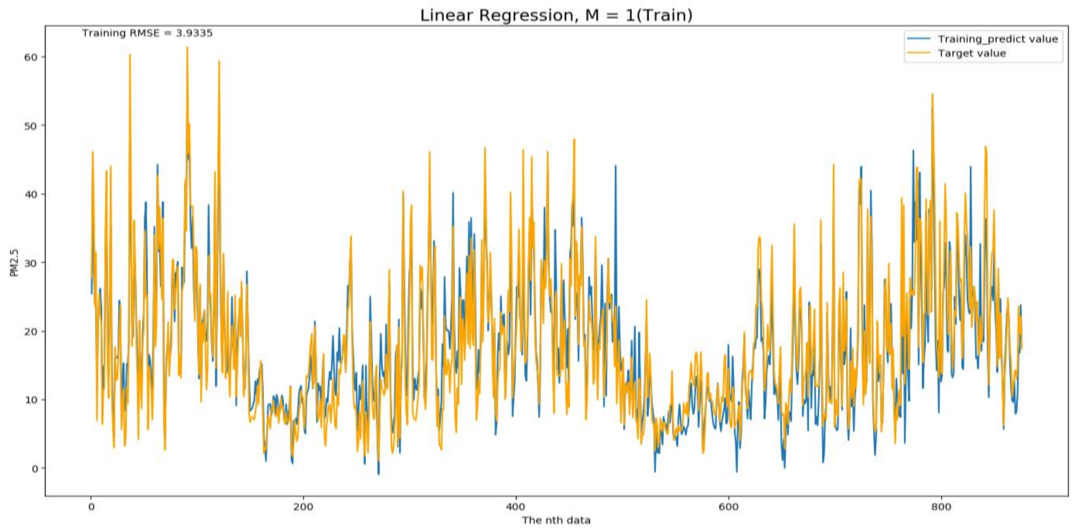

Figure 1 : M=1 時，模型對訓練數據集的預測及實際值的比較與 RMSE。

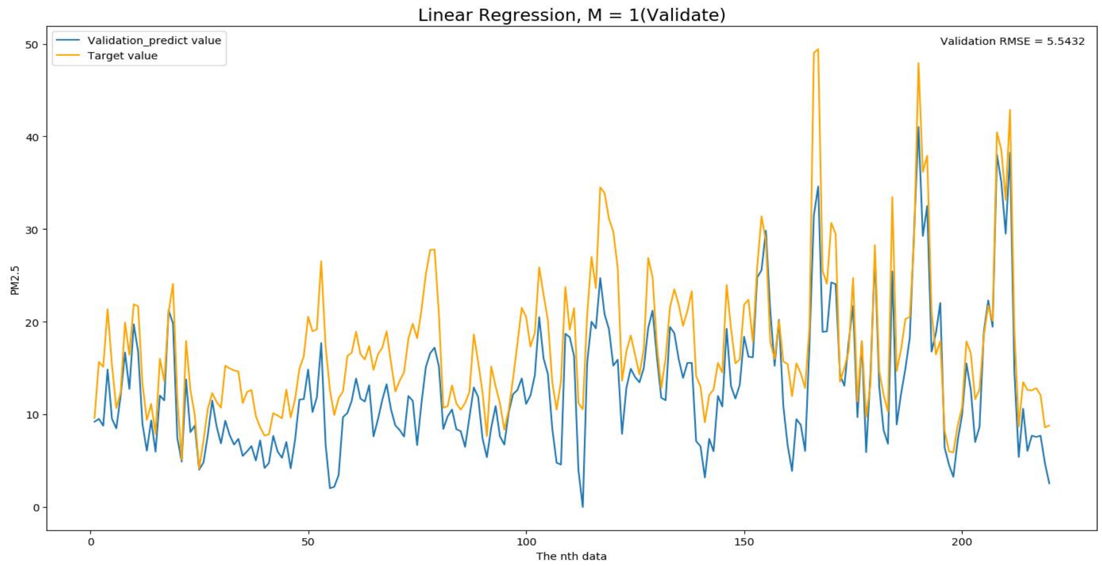

Figure 2 : M=1 時，模型對驗證數據集的預測及實際值的比較與 RMSE。

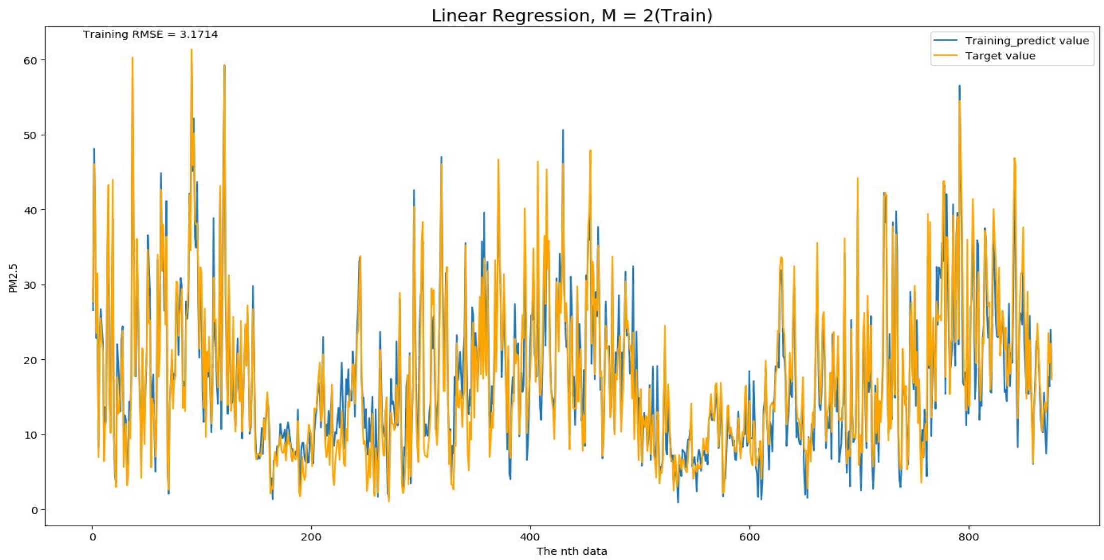

Figure 3 : M=2 時，模型對訓練數據集的預測及實際值的比較與 RMSE。

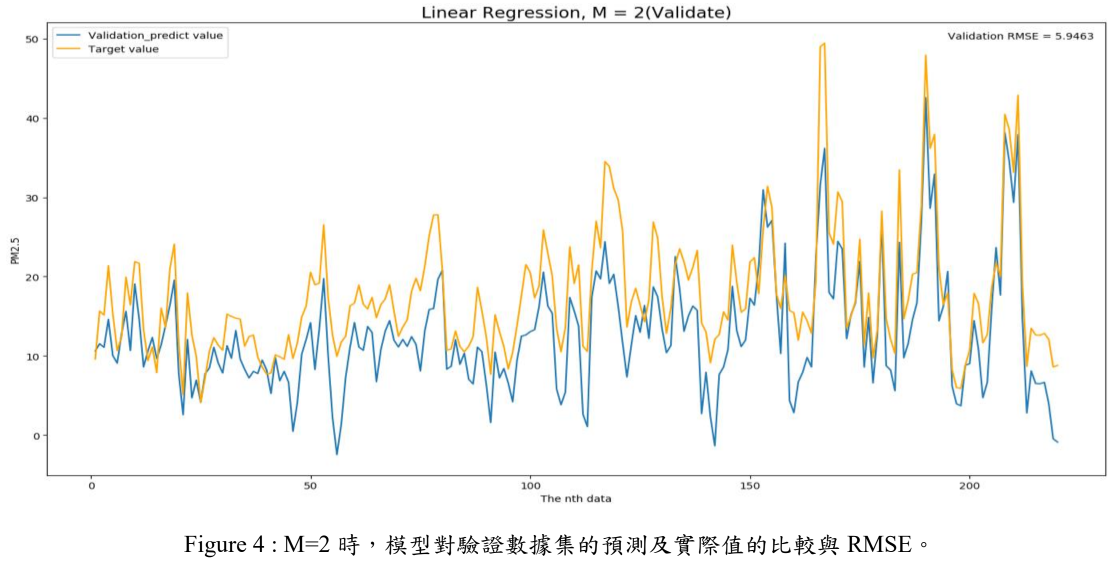

Figure 4 : M=2 時，模型對驗證數據集的預測及實際值的比較與 RMSE。

	M = 1	M = 2
Training	3.9335	3.1769
Validation	5.5432	5.9463

Table 1：不同模型下，在訓練數據集與驗證數據集的 RMSE。

結果討論：

從 Figure 1 到 Figure 4 與 Table 1 可以觀察到在訓練數據集中，二次多項式迴歸的模型有較低的損失，然而到了驗證數據集中，二次多項式迴歸的損失卻反過來超越了一次線性迴歸在驗證數據集上的損失。發生此現象的原因推測是源於模型的複雜度所造成，在一次線性迴歸模型中，權重函數為一個(18,1)的向量(常數項+特徵數目(D))，而在二次多項式迴歸模型中，權重函數為一個(171,1)的向量(常數項+特徵數目+(特徵數目*(特徵數目+1)/2))，因此對於此次問題的數據集來說，二次多項式迴歸模型或許太過複雜，導致模型的變異數大(對陌生資料較為敏感)，偏差小(在訓練數據集的準確度較高)。因此在面對驗證數據集時，得到較大的損失，反觀一次線性迴歸模型，雖然偏差較大(在訓練數據集上比較不準)，但有較小的變異數(對陌生資料適應性較好)，所以在測試數據集與驗證數據集的表現上不至於相差太多。藉由調整測試數據集中的樣本數目(如 Figure 5 所示)，可以發現二次多項式迴歸模型由於複雜度的關係，更容易受到訓練數據多寡的影響。

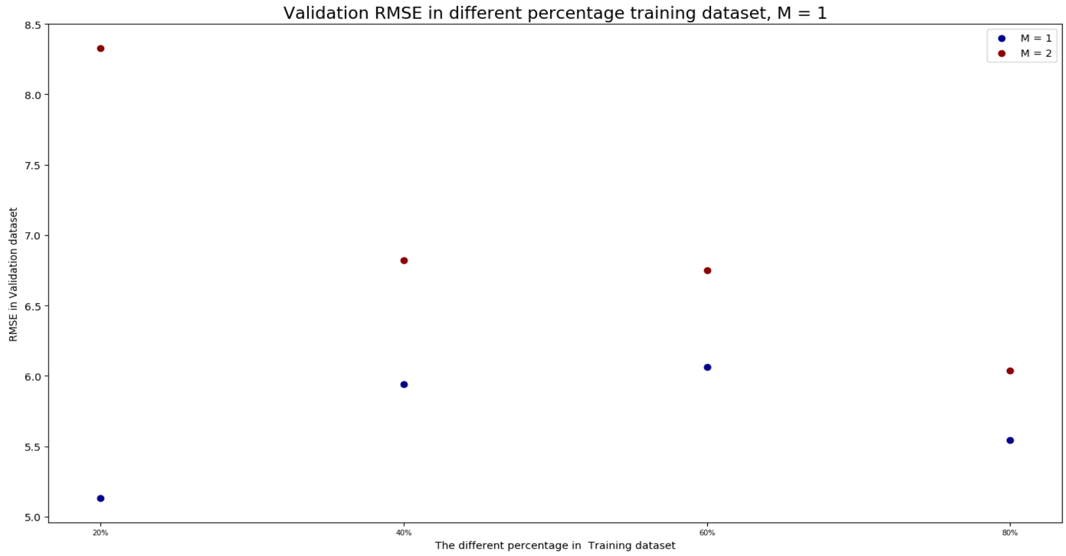

Figure 5：在不同比例的測試數據集訓練後，在驗證數據集得到的 RMSE。

(b)小題

討論與結果呈現：

在(a)小題提到模型複雜度的影響，而影響複雜度的一大因素就是所使用的特徵數目(D)，如果所選用的特徵越少，模型的複雜度就會越低，如果能排除不必要的特徵，將有助於減緩過擬合的發生。因此我們可以藉由分析線性迴歸模型的權重值來選擇我們所需要的特徵。一階線性方程在訓練數據集的 RMSE 為 3.9335，各特徵權重如 Figure 6 所示。緊接著為了觀察每個特徵對結果產生的影響，我們每次分別移除一個不同的特徵，並與初始結果進行比較，來觀察哪些特徵會對結果產生較大的影響，測試結果如 Figure 7 所示。可以發現 PM10 與 CO 兩個特徵會對結果產生較大的影響，Rainfall, RH, WD_HR, WIND_DIREC 也會對結果產生一定的影響，而其他的特徵則對結果幾乎沒有影響。因此我認為其他沒有提到的特徵對於

預測 PM2.5 來說是相對不重要的，如果模型面臨複雜度偏高的情形(例如過擬合)，我們就可以考慮移除這些對結果影響不大的特徵，來降低模型的複雜度，減緩過擬合的發生。在下一題的討論中，就會藉由移除不必要的特徵，來避免模型產生過擬合的情形。

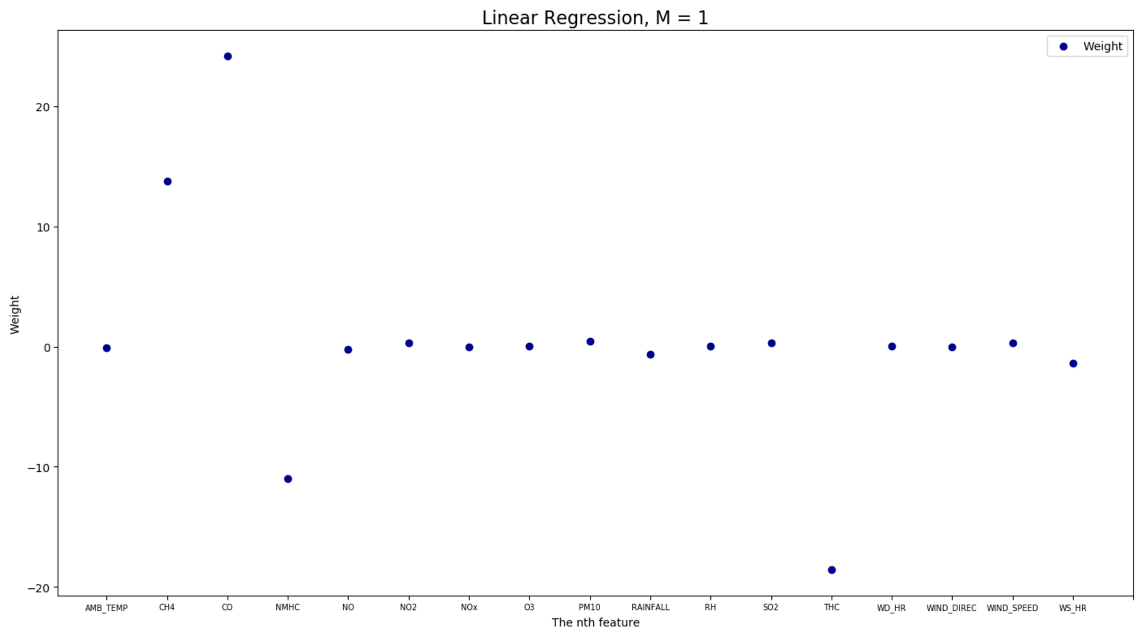

Figure 6 各個特徵所對應的權重值

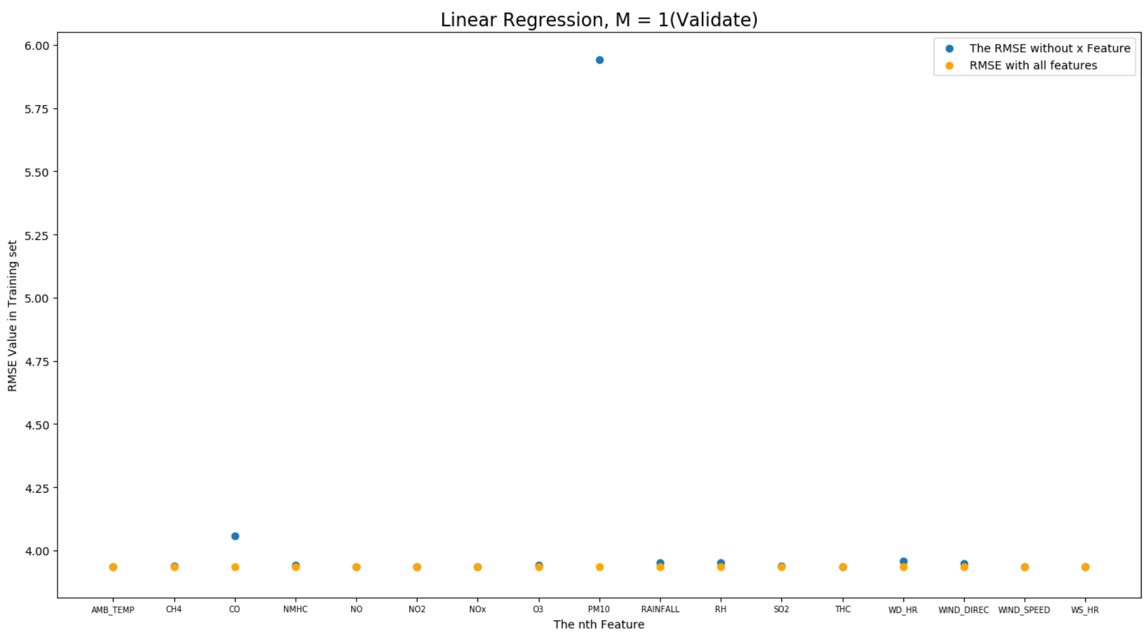

Figure 7 移除各特徵權重對 RMSE 的影響

2. Maximum likelihood approach

(a)小題

模型設置：

	Model 1	Model 2	Model 3
Basic Function	Polynomial (M = 2)	Gaussian	Sigmoidal

模型共有設置：

使用特徵維度(D)	17
訓練資料佔比	70%
Error Function	RMSE

實做結果：

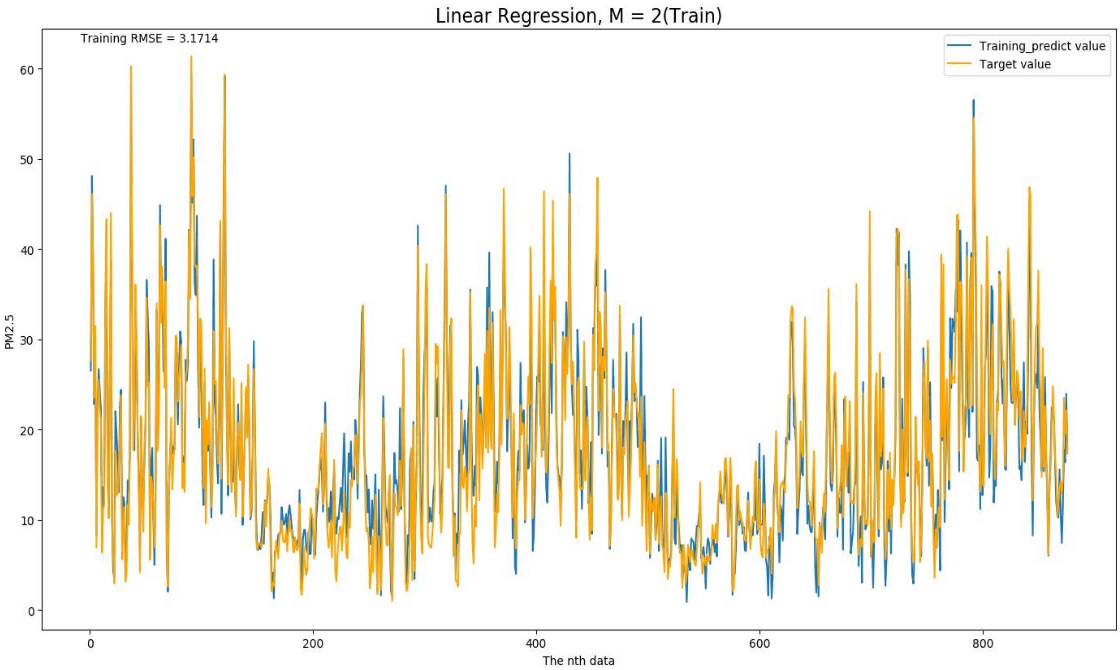

Figure 8 : M=2 時，模型對訓練數據集的預測及實際值的比較與 RMSE。

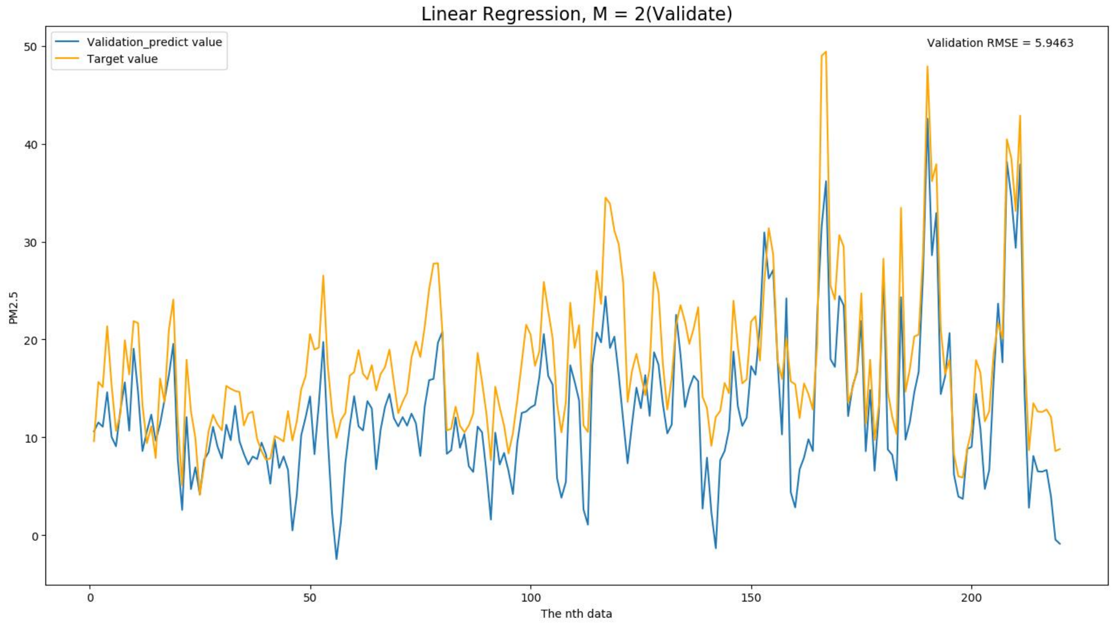

Figure 9 : M=2 時，模型對驗證數據集的預測及實際值的比較與 RMSE。

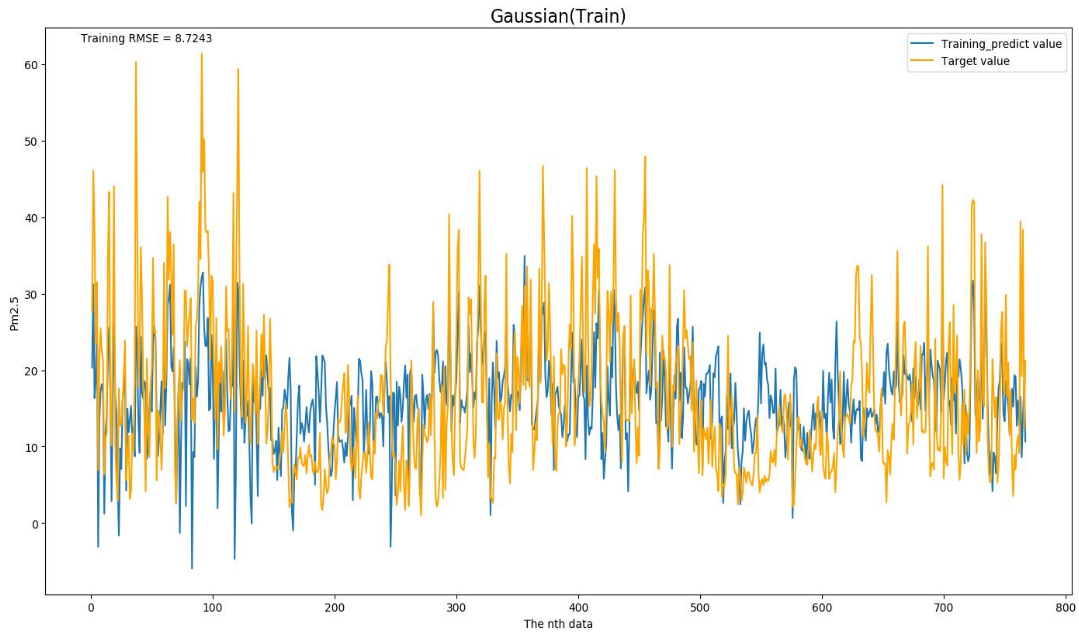

Figure 10 : Gaussian 做為 Basic Function 時，模型對訓練數據集的預測及實際值的比較與 RMSE。

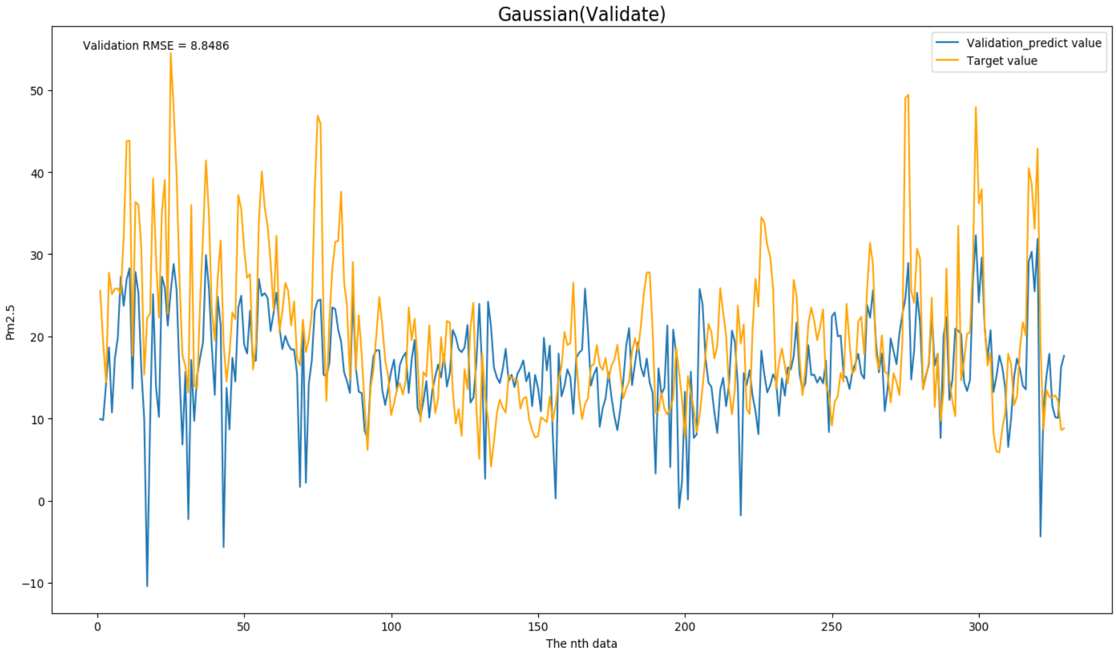

Figure 11 : Gaussian 做為 Basic Function 時，模型對驗證數據集的預測及實際值的比較與 RMSE。

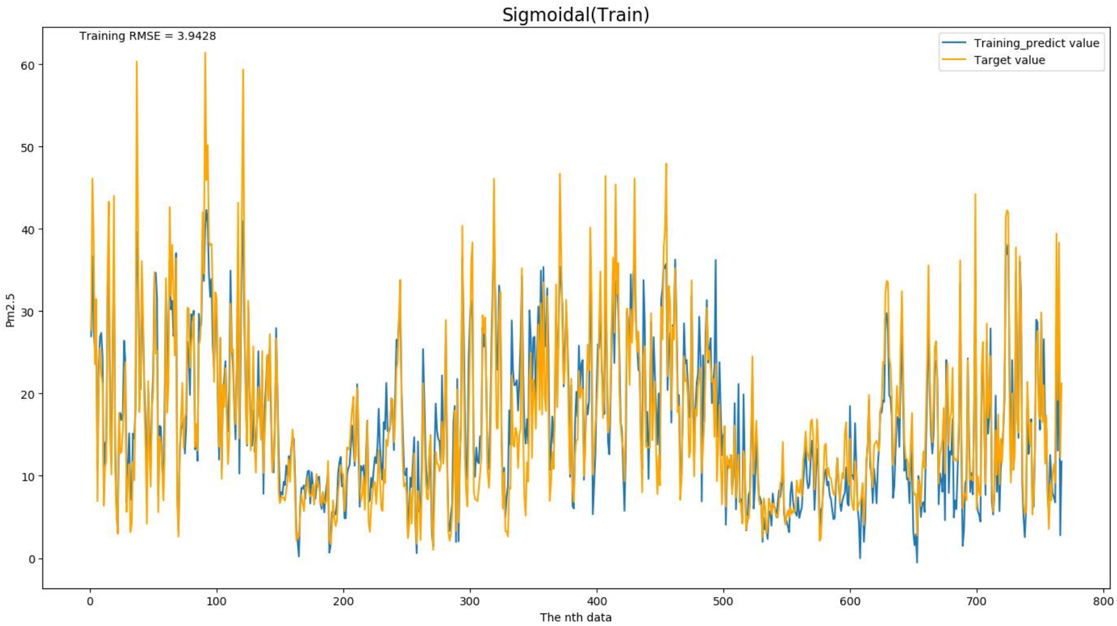

Figure 12 : Sigmoidal 做為 Basic Function 時，模型對訓練數據集的預測及實際值的比較與 RMSE。

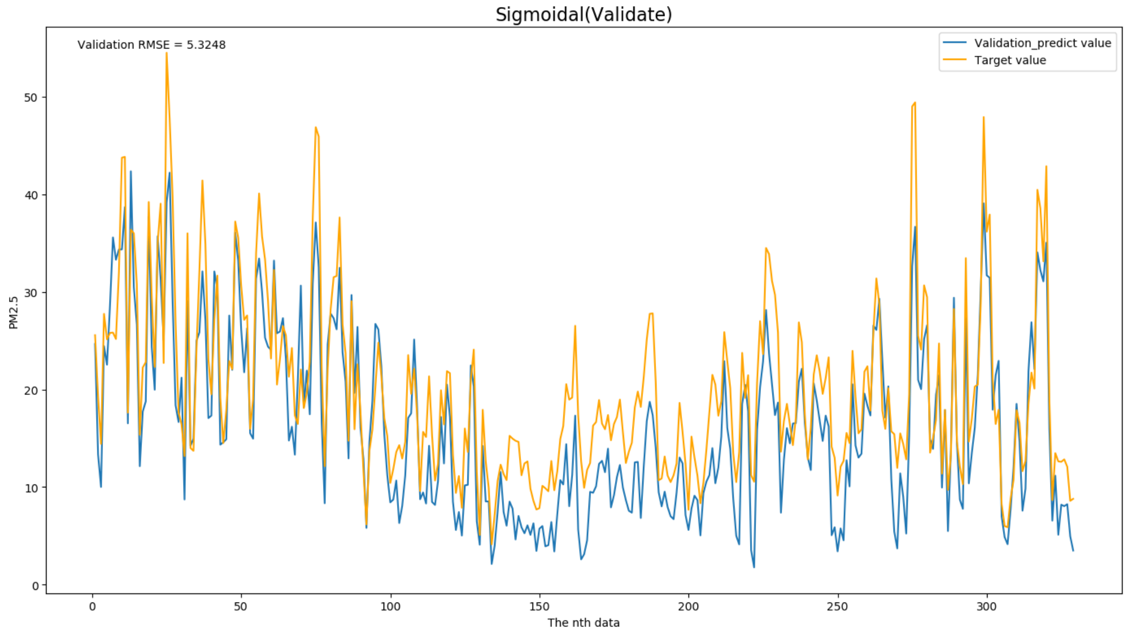

Figure 13 : Sigmoidal 做為 Basic Function 時，模型對驗證數據集的預測及實際值的比較與 RMSE。

	Polynomial(M=2)	Gaussian	Sigmoidal
Training	3.1714	8.7243	3.9428
Validation	5.9463	8.8486	5.3248

Table 2： 不同模型下，在訓練數據集與驗證數據集的 RMSE。

討論 ：

從 Table 2 中可以發現，使用二次多項式迴歸以及 Gaussian 的模型都產生了過度擬合的現象(在訓練數據集與驗證數據集的表現差距很大或是無法很好的收斂)，這是由於這兩種 Basic Function 的複雜度過高，在特徵數目過多的情況下發生了過擬合的現象。要解決此現象可以透過調整模型的超參數:一個是改變模型本身的複雜度，例如將二次多項式迴歸轉為一次線性迴歸，而此方法我們在 2-1 題的討論中就已經得到驗證，一階線性迴歸的模型並不會產生過擬合的

現象。另一個則是調整輸入的特徵維度，藉由減少不必要的輸入，就可以降低整體模型的複雜度，我們在(b)小題中，我們將引入交叉驗證的方式並藉由調整輸入特徵維度的方式來嘗試解決過擬合的問題。

(b)小題

事前討論：我們需要選擇要刪除的特徵維度，因此我們可以參考 Figure 6 與 Figure 7 的結果，最後我選擇出了兩種組合方式，第一組是只保留影響最大的五項特徵，也就是只保留 PM10, CO, Rainfall, RH, WD_HR 這五項特徵(D=5);第二組選擇法則是刪除影響最小的五項特徵，也就是刪除 AMB_TEMP, NO, NO2, NOx, THC 這五項特徵(D=12)，並將這結果與不進行特徵篩選的結果進行比對。

模型設置：

	Model 1	Model 2	Model 3
Basic Function	Polynomial (M = 2)	Gaussian	Sigmoidal

模型共有設置：

N-Fold	N = 9
Error Function	RMSE

實做結果：

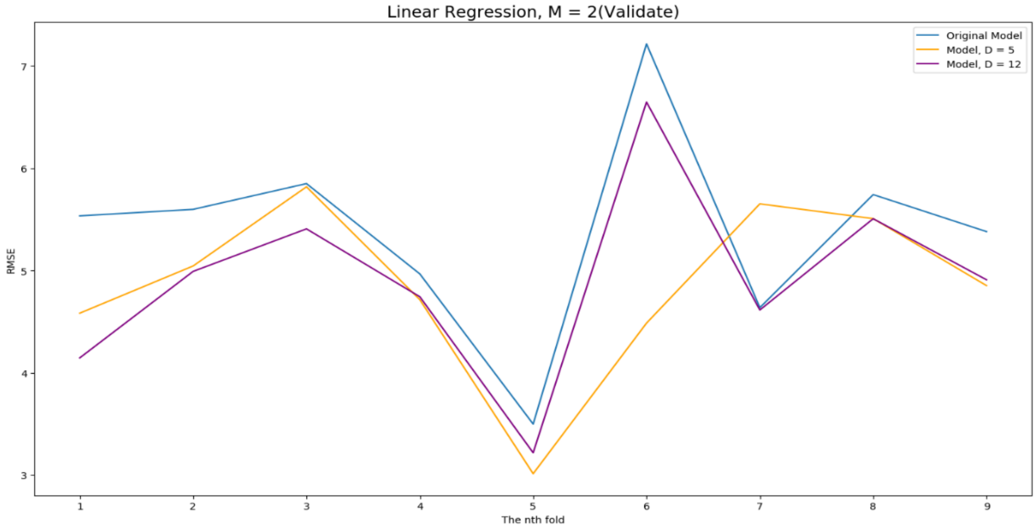

Figure 14 在二次線性迴歸模型減少輸入的特徵維度

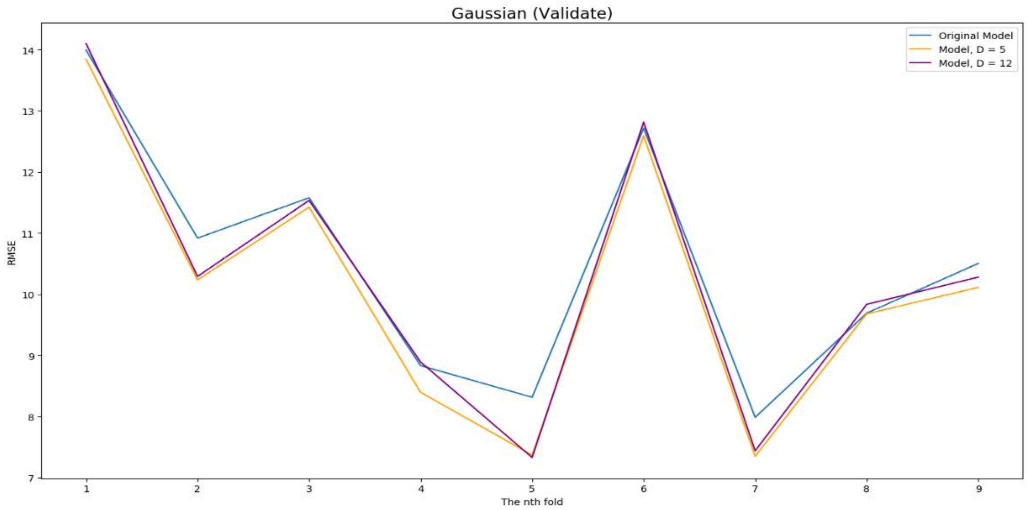

Figure 15 在 Gaussian 模型減少輸入的特徵維度

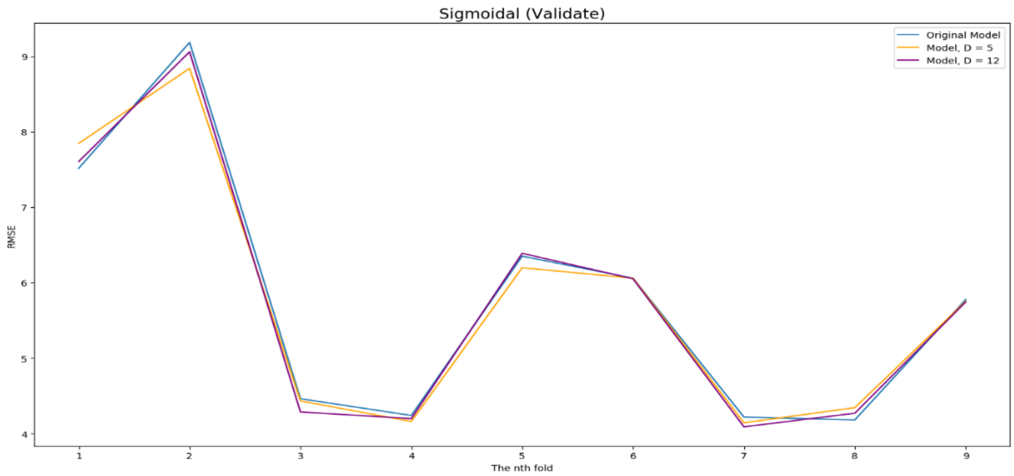

Figure 16 在 Sigmoidal 模型減少輸入的特徵維度

討論 :

在減少輸入的特徵維度之後，二階多項式迴歸模型與高斯模型的預測效能皆有得到一定的改善，而 Sigmoidal 模型則沒有受到太多的影響(因為原本就沒有受到過擬合的影響)。從結果可以確認減少輸入的特徵維度，的確可以幫助減少模型的複雜度，並改善過擬合的問題。

3. Maximum a posteriori approach

(a)小題

模型設置：

Basic Function	Gaussian
權重計算方式	Maximum a posteriori approach
訓練資料佔比	70%
Error Function	RMSE
Alpha	5
Beta	10

實做結果：

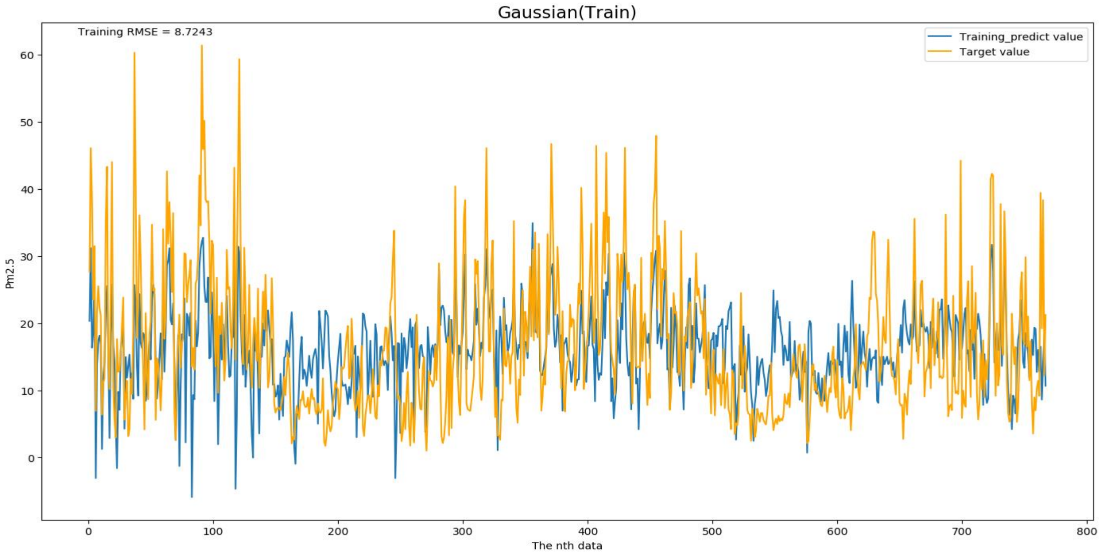

Figure 17 : Gaussian 模型使用 W_{map} 時，對訓練數據集的預測及實際值的比較與 RMSE。

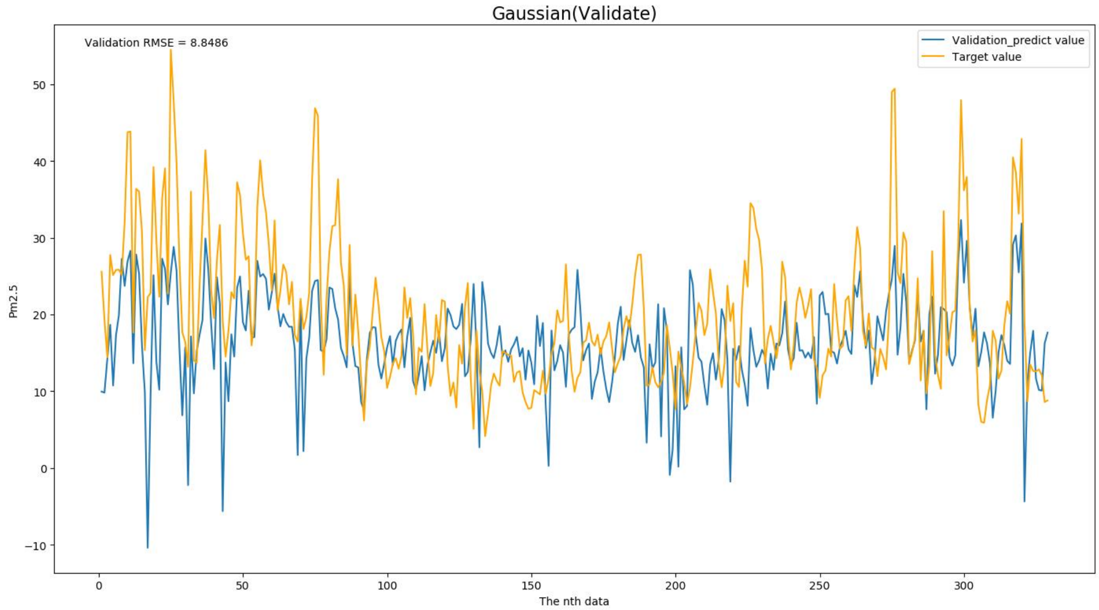

Figure 18 : Gaussian 模型使用 W_{map} 時，對訓練數據集的預測及實際值的比較與 RMSE。

	W_{ML}	W_{map}
Training	8.7243	8.7243
Validation	8.8486	8.8486

Table 3 : Gaussian 模型分別使用 W_{ML} 與 W_{map} 時，所得到的 RMSE

討論 :

從 Table 3 中可以發現，在不調整超參數的情況下，不管是使用 W_{ML} 或是 W_{map} 的 Gaussian 模型都產生了過度擬合的現象，這表示即使改變了權重的計算方式，對於這問題而言，此模型還是太過複雜了。因此我們需要再次減少輸入的特徵維度。

(b)小題

事前討論 : 我們此次只保留影響最大的五項特徵，也就是只保

留 PM10, CO, Rainfall, RH, WD_HR 這五項特徵(D=5)，並將這結果與不進行特徵篩選的結果進行比對。

實做結果：

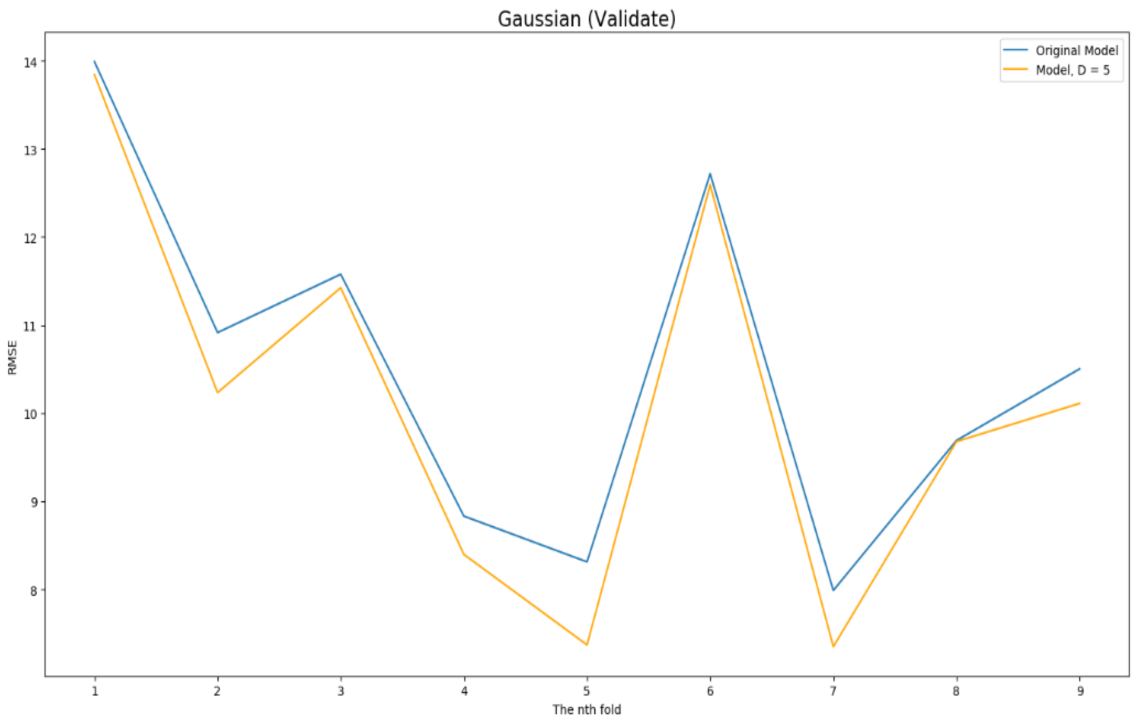

Figure 19：在使用 W_{map} 的 Gaussian 模型減少輸入的特徵維度

討論：

可以看到減少輸入特徵維度之後，效能一樣有更加進步的趨勢，且使用 W_{map} 的模型在表現上更是比使用 W_{ML} 的模型來的更好一些。

(c)小題

討論：

Maximum a Posteriori approach(MAP)相比於Maximum Likelihood(ML)就好像是在W_{ML}上加入了Regularization一樣(第一題

的證明即表明了此事)。因此W_{map}相比於W_{ML}，可以更好的解決過擬合的問題。可以透過Figure 15與Figure 19證明此事。