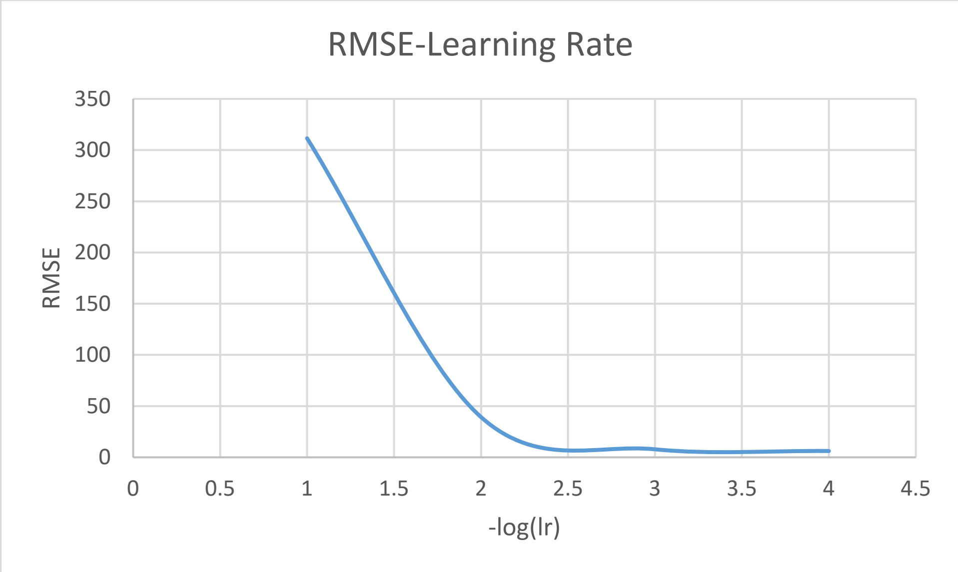
Homework 1 Report - PM2.5 Prediction

學號： B05901082 系級： 電機三 姓名: 楊晟甫

1. (1%) 請分別使用至少 **4** 種不同數值的 learning rate 進行 training（其他參數需一致），對其作圖，並且討論其收斂過程差異。



lr	RMSE
0.1	311.514
0.01	39.31
0.001	7.85
0.0001	6.17

這些都是在 iteration = 200 的情況下所觀測的，可以發現從 1e-3 開始 RMSE 就開始收斂，所以我最後選擇 1e-4 當作我的 learning rate

在 training 過程中可以明顯觀察到，在 learning rate 值大的時候 RMSE 改變的幅度很大，因此容易發生老師在上課時所提及的走不到最佳解的問題，也就是一次改變太多不小心”走過頭”，因此將 learning rate 調小可以有效解決逼近的準確度。

2. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項（含 bias 項）以及每筆 data9 小時內 PM2.5 的一次項（含 bias 項）進行 training，比較並討論這兩種模型的 root mean-square error（根據 kaggle 上的 public/private score）。

	Public	**Private**	**Train RMS**
PM2.5	7.62095	7.77712	6.18243
All	7.21541	7.52242	6.11485

其實兩者的差距並不顯著，若以常理來推斷，雨量、PM10 都有可能對 PM2.5 的值造成影響，但是由此可發現前 9 小時 PM2.5 的值對於接下來的第 10 小時的值的影響是很重大的，因此可能可以自行產生 PM2.5 的平方，做為新的 feature，來增加 PM2.5 在所有 feature 一起 train 時的影響力，讓 model 準確性更高。

3. (1%)請分別使用至少四種不同數值的 regulization parameter λ 進行 training（其他參數需一至），討論及討論其 RMSE(traning, testing)（testing 根據 kaggle 上的 public/private score）以及參數 weight 的 L2 norm。

Lambda	*Training RMS*	*L2 norm*	*Public score*	*Private Score*
0.0001	6.2354	0.7005	7.30413	7.58808
0.001	6.2075	0.7020	7.31667	7.35577
0.01	6.3607	0.7044	7.56700	7.41874
0.1	6.6458	0.7442	8.55031	7.78667
1	6.6797	0.7567	9.47602	7.72320
10	6.7745	0.7601	9.74538	7.62044
0(no regularization)	6.2004	0.7003	7.42463	7.35639

可以發現隨著 lambda 的減少，training 的 RMS 也減少，0.001 的時候達到最低，public 跟 private 也隨之進步，但是當我們與未做 regularization 的數據比較時，發現 lambda 太大反而造成誤差提升，因此歸納的結果顯示只有當找到最佳的 lambda 值時，regularization 才能對此 model 產生助益。

4~6 (3%) 請參考數學題目（連結：），將作答過程以各種形式（latex 尤佳）清楚地呈現在 pdf 檔中（手寫再拍照也可以，但請注意解析度）
詳見下頁附圖

No:

Date: /

HW1 電機三 B05901082 楊晟甫

4. (a) $\frac{\partial}{\partial w} E_D(w) = -\sum_{n=1}^{N} r_n \{t_n - w^T x_n\} x_n = 0 \quad \sum_{n=1}^{N} r_n t_n x_n \left(\sum_{n=1}^{N} r_n x_n x_n^T\right) w$

$w^* = \left(\sum_{n=1}^{N} r_n x_n x_n^T\right)^{-1} \left(\sum_{n=1}^{N} r_n t_n x_n\right)$

(b) $w^* = \left(2 \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot [2\ 3] + 1 \cdot \begin{bmatrix} 5 \\ 1 \end{bmatrix} \cdot [5\ 1] + 3 \begin{bmatrix} 5 \\ 6 \end{bmatrix} [5\ 6]\right)^{-1} \cdot \left(10 \begin{bmatrix} 5 \\ 1 \end{bmatrix} + 15 \begin{bmatrix} 5 \\ 6 \end{bmatrix}\right)$

$= \begin{bmatrix} 2.2827 \\ -1.1358 \end{bmatrix}$

5. $E'_D(w) = \frac{1}{2} \sum_{n=1}^{N} \left(y(x_n, w) - t_n + \sum_{d=1}^{D} w_d \varepsilon_{nd}\right)^2 = \frac{1}{2} \sum_{n=1}^{N} \Big\{ (y(x_n, w) - t_n)^2 + 2(y(x_n, w) - t_n)$
$\left(\sum_{d=1}^{D} w_d \varepsilon_{nd}\right) + \left(\sum_{d=1}^{D} w_d \varepsilon_{nd}\right)^2 \Big\}$

$E[E'_D(w)] = \frac{1}{2} \sum_{n=1}^{N} \left\{ (y(x_n, w) - t_n)^2 + 2(y(x_n, w) - t_n)\left(\sum_{d=1}^{D} w_d \underbrace{E[\varepsilon_{nd}]}_{0}\right) + E\left[\left(\sum_{d=1}^{D} w_d \varepsilon_{nd}\right)^2\right] \right\}$

$E\left[\left(\sum_{d=1}^{D} w_d \varepsilon_{nd}\right)^2\right] = \sum_{d=1}^{D} \sum_{d'=1}^{D} w_d w_{d'} E[\varepsilon_{nd} \varepsilon_{nd'}] = \sum_{d=1}^{D} \sum_{d'=1}^{D} w_d w_{d'} \delta_{dd'} = \sum_{d=1}^{D} w_d^2$

$\therefore E[E'_D(w)] = \frac{1}{2} \sum_{n=1}^{N} \left\{ (y(x_n, w) - t_n)^2 + \sum_{d=1}^{D} w_d^2 \right\} = E_D(w) + \frac{N}{2} \sum_{d=1}^{D} w_d^2$ #

6. $A = PDP^{-1} \quad \det|A| = \det|P| \cdot \det|D| \cdot \det|P^{-1}| = 1 \cdot \prod_{i=1}^{n} \lambda_i \cdot 1 = \prod_{i=1}^{n} \lambda_i$

$A \in R^{n \times n} \quad \text{trace}(A) = \sum_{i=1}^{n} \lambda_i \quad D$: diagnose

$\frac{d}{d\alpha} \ln|A| = \frac{d \ln|A|}{d\lambda} \cdot \frac{d\lambda}{d\alpha} = \frac{1}{\lambda_1} \cdot \frac{d\lambda_1}{d\alpha} + \cdots + \frac{1}{\lambda_n} \cdot \frac{d\lambda_n}{d\alpha} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \frac{d\lambda_i}{d\alpha}$

$\text{Trace}\left(A^{-1} \frac{dA}{d\alpha}\right) = \text{Trace}\left(PD^{-1}P^{-1} \cdot \frac{d(PDP^{-1})}{d\alpha}\right) = \text{Trace}\left(D^{-1} \cdot \frac{dA}{d\alpha}\right)$

$\frac{dA}{d\alpha} = \frac{\partial(P^{-1}DP)}{\partial \lambda_1} \frac{d\lambda_1}{d\alpha} + \cdots + \frac{\partial(P^{-1}DP)}{\partial \lambda_n} \cdot \frac{d\lambda_n}{d\alpha} = \sum_{i=1}^{n} P^{-1} B_i P \cdot \frac{d\lambda_i}{d\alpha} \quad B_i \Rightarrow \begin{matrix} (i,i) = 1 \\ \text{o.w.} = 0 \end{matrix}$

$= \text{Trace}\left(D^{-1} \cdot \sum_{i=1}^{n} P^{-1} B_i P \cdot \frac{d\lambda_i}{d\alpha}\right) = \sum_{i=1}^{n} \text{Trace}\left(D^{-1} P^{-1} B_i P \cdot \frac{d\lambda_i}{d\alpha}\right) = \sum_{i=1}^{n} \text{Trace}\left(D^{-1} B_i \frac{d\lambda_i}{d\alpha}\right)$

$= \sum_{i=1}^{n} \frac{1}{\lambda_i} \frac{d\lambda_i}{d\alpha}$ # 故左式 = 右式 #