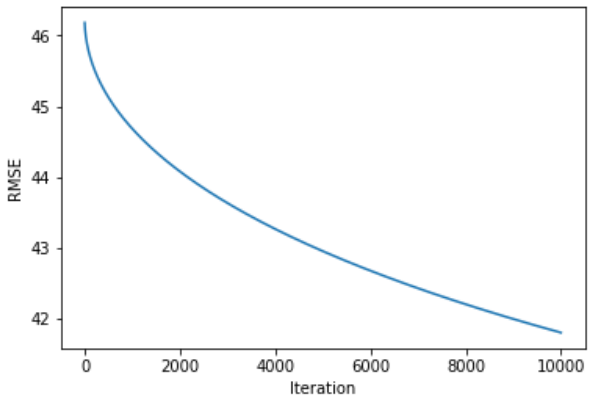
Homework 1 Report - PM2.5 Prediction

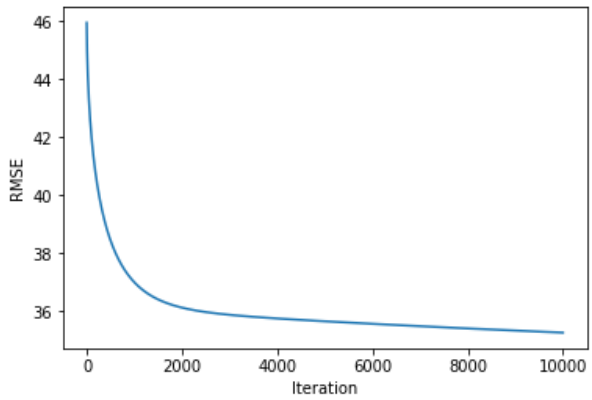
學號：R07922142　系級：資工所碩一　姓名:歐政鷹

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。

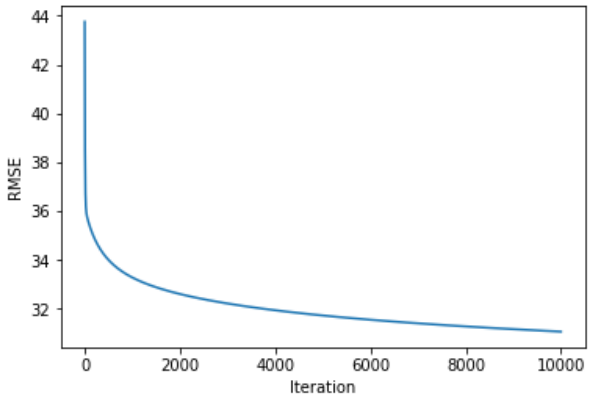
以下為示意圖：



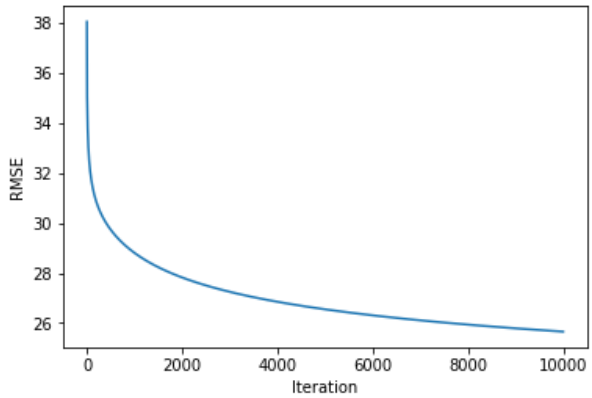
learning rate：0.001



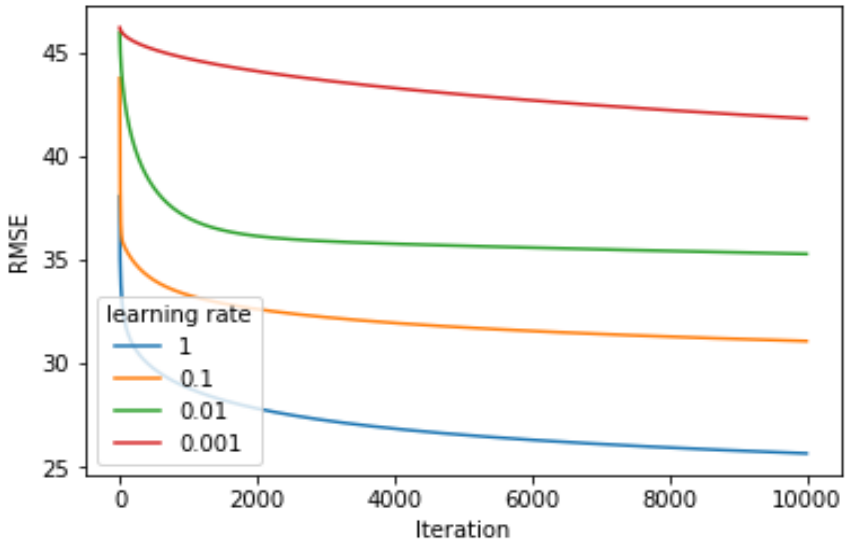
learning rate：0.01



learning rate：0.1



learning rate：1



以上表格中前面四張圖表分別代表了利用不同 learning rate 作訓練時 RMSE 的變化情況，最後一張圖表則是同時針對四種情況作比較。

可以看出當 learning rate 越大時，RMSE 在前 2000 次的下降速度就會越快，然後會保持在一定的速度下下降，若想要快速地降低訓練時的 Error，我們可以用較大的 learning rate，但其實同時也有機會衍生出 over fitting 的情況。

所以在訓練時我們應多嘗試不同的 learning rate，並利用 validation set 來檢查怎樣的 learning rate 可以讓我們不會出現 over fitting 的情況而有可以有好的 testing 結果。

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）。

Testing error	Public score	Private score
All feature	7.56222	7.64842
Only PM2.5	10.89036	10.35148

從以上表格可以很明顯的看出只使用 PM2.5 的資料作訓練時的 Testing error 比起用所有資料作訓練的 Testing error 大很多。

原因很明顯是因為只用 PM2.5 作訓練的話會有很多可以反映出 PM2.5 數值的數據沒有被考慮到，例如 PM10、風速、風量等等。換句話說就是訓練數據太過平滑，很多應該考慮進去的資料都被丟掉了，訓練出一個不好的模型，導致 Testing error 很大。

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。

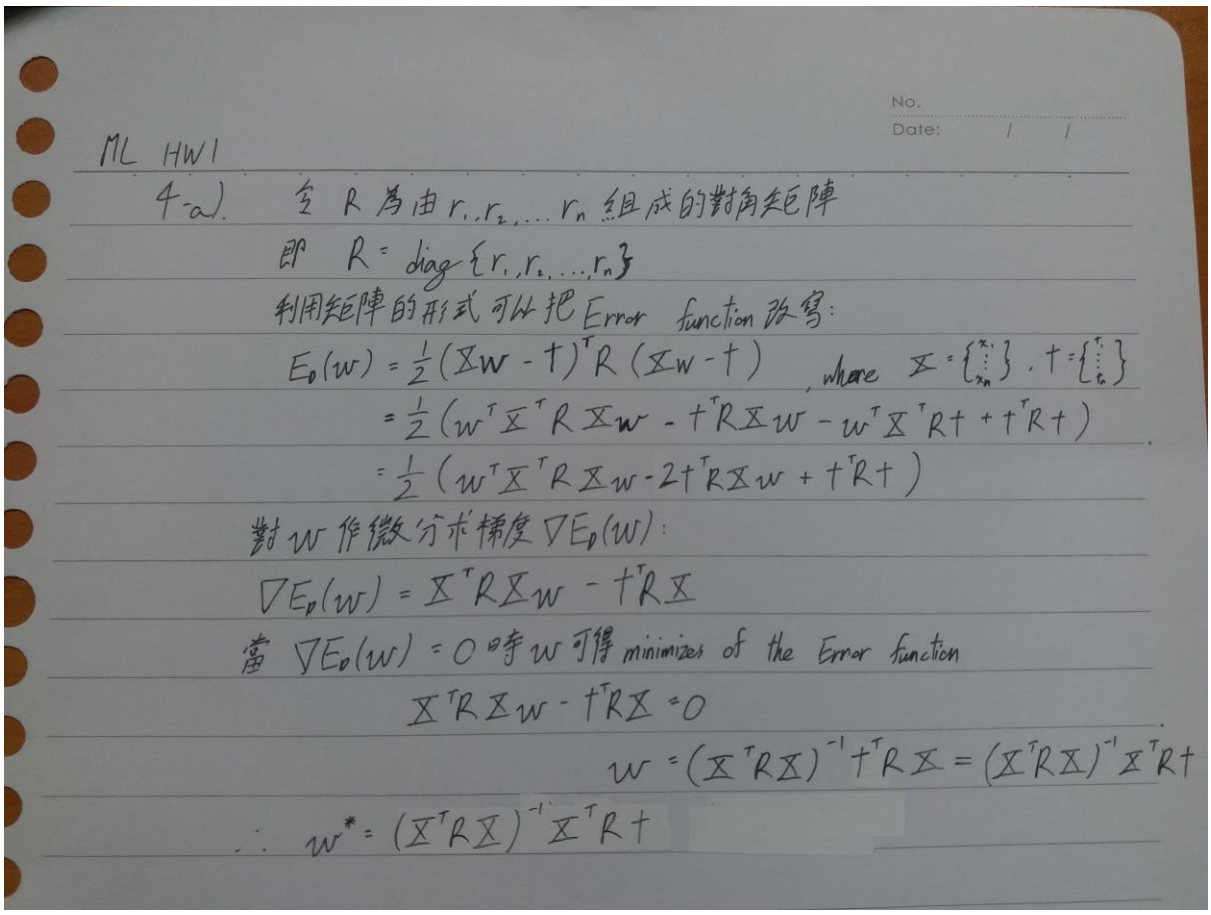
Training λ	Public score	Private score	L2 norm
0	8.17091	8.16487	1.0293574368444218

0.1	7.54318	7.63110	0.9584548894901731
1	7.56222	7.64842	0.8178637160782148
10	7.73218	7.71069	0.6688596761405721

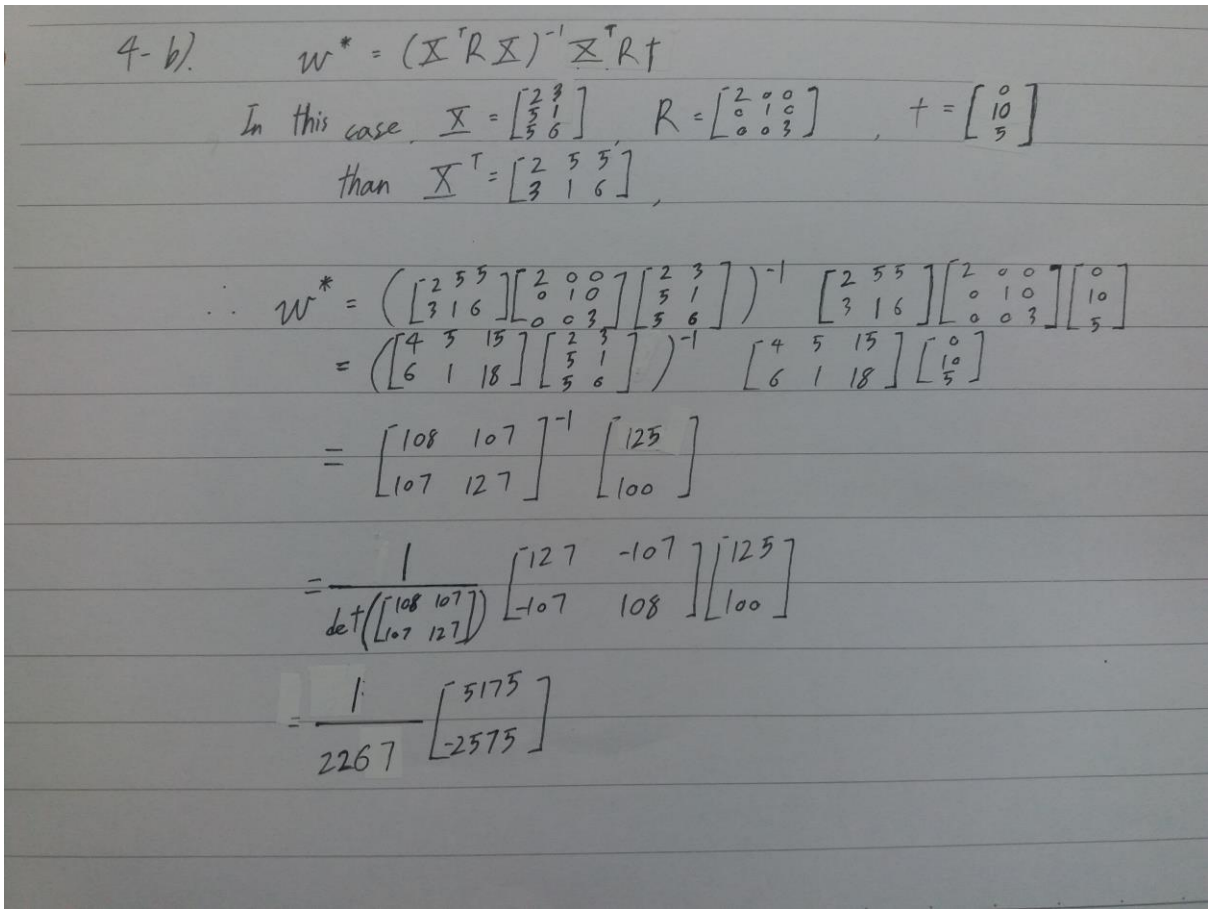
由上表可知，當 regulization parameter λ 越大時，testing error 的錯誤率也會慢慢提高。但是 λ= 0 時的錯誤率比 λ= 0.1 時的大，證明我們的資料中還是含有一些雜訊，不能無視。

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）。

(4-a)



(4-b)


4-b). $w^* = (X^T R X)^{-1} X^T R t$
In this case, $X = \begin{bmatrix} 2 & 3 \\ 3 & 1 \\ 5 & 6 \end{bmatrix}$, $R = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, $t = \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}$
than $X^T = \begin{bmatrix} 2 & 3 & 5 \\ 3 & 1 & 6 \end{bmatrix}$,
$\therefore w^* = \left(\begin{bmatrix} 2 & 3 & 5 \\ 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 1 \\ 5 & 6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & 3 & 5 \\ 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}$
$= \left(\begin{bmatrix} 4 & 3 & 15 \\ 6 & 1 & 18 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 1 \\ 5 & 6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 4 & 3 & 15 \\ 6 & 1 & 18 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}$
$= \begin{bmatrix} 108 & 107 \\ 107 & 127 \end{bmatrix}^{-1} \begin{bmatrix} 125 \\ 100 \end{bmatrix}$
$= \frac{1}{\det\left(\begin{bmatrix} 108 & 107 \\ 107 & 127 \end{bmatrix} \right)} \begin{bmatrix} 127 & -107 \\ -107 & 108 \end{bmatrix} \begin{bmatrix} 125 \\ 100 \end{bmatrix}$
$= \frac{1}{2267} \begin{bmatrix} 5175 \\ -2575 \end{bmatrix}$

5

ML HW1

5. 4/2

$$\hat{y}_n = w_0 + \sum_{i=1}^{D} w_i (x_{i,n} + \epsilon_i)$$
$$= w_0 + \sum_{i=1}^{D} w_i x_{i,n} + \sum_{i=1}^{D} w_i \epsilon_i$$
$$= y(x_n, w) + \sum_{i=1}^{D} w_i \epsilon_i$$
$$= y_n + \sum_{i=1}^{D} w_i \epsilon_i \quad , \text{let } y_n = y(x_n, w)$$

$$\therefore \hat{E}(w) = \frac{1}{2}\sum_{n=1}^{N} (\hat{y}_n - t_n)^2$$
$$= \frac{1}{2}\sum_{n=1}^{N} \left(\hat{y}_n^2 - 2\hat{y}_n t_n + t_n^2\right)$$
$$= \frac{1}{2}\sum_{n=1}^{N} \left(\left(y_n + \sum_{i=1}^{D} w_i \epsilon_i\right)^2 - 2\left(y_n + \sum_{i=1}^{D} w_i \epsilon_i\right) t_n + t_n^2 \right)$$
$$= \frac{1}{2}\sum_{n=1}^{N} \left[y_n^2 + 2y_n \sum_{i=1}^{D} w_i \epsilon_i + \left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2 - 2t_n y_n - 2t_n \sum_{i=1}^{D} w_i \epsilon_i + t_n^2 \right]$$
$$= \frac{1}{2}\sum_{n=1}^{N} \left[y_n^2 - 2t_n y_n + t_n^2 + 2y_n \sum_{i=1}^{D} w_i \epsilon_i - 2t_n \sum_{i=1}^{D} w_i \epsilon_i + \left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2 \right]$$

Consider $E_x[\hat{E}(w)]$, $E(w)$的期望值

$$E_x[\hat{E}(w)] = E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left[y_n^2 - 2t_n y_n + t_n^2\right] + \frac{1}{2}\sum_{n=1}^{N}\left[2y_n \sum_{i=1}^{D} w_i \epsilon_i - 2t_n \sum_{i=1}^{D} w_i \epsilon_i + \left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2\right]\right]$$
$$= E_x[E(w)] + E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left[2y_n \sum_{i=1}^{D} w_i \epsilon_i - 2t_n \sum_{i=1}^{D} w_i \epsilon_i + \left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2\right]\right]$$

$\because E_x[\epsilon_i] = 0$

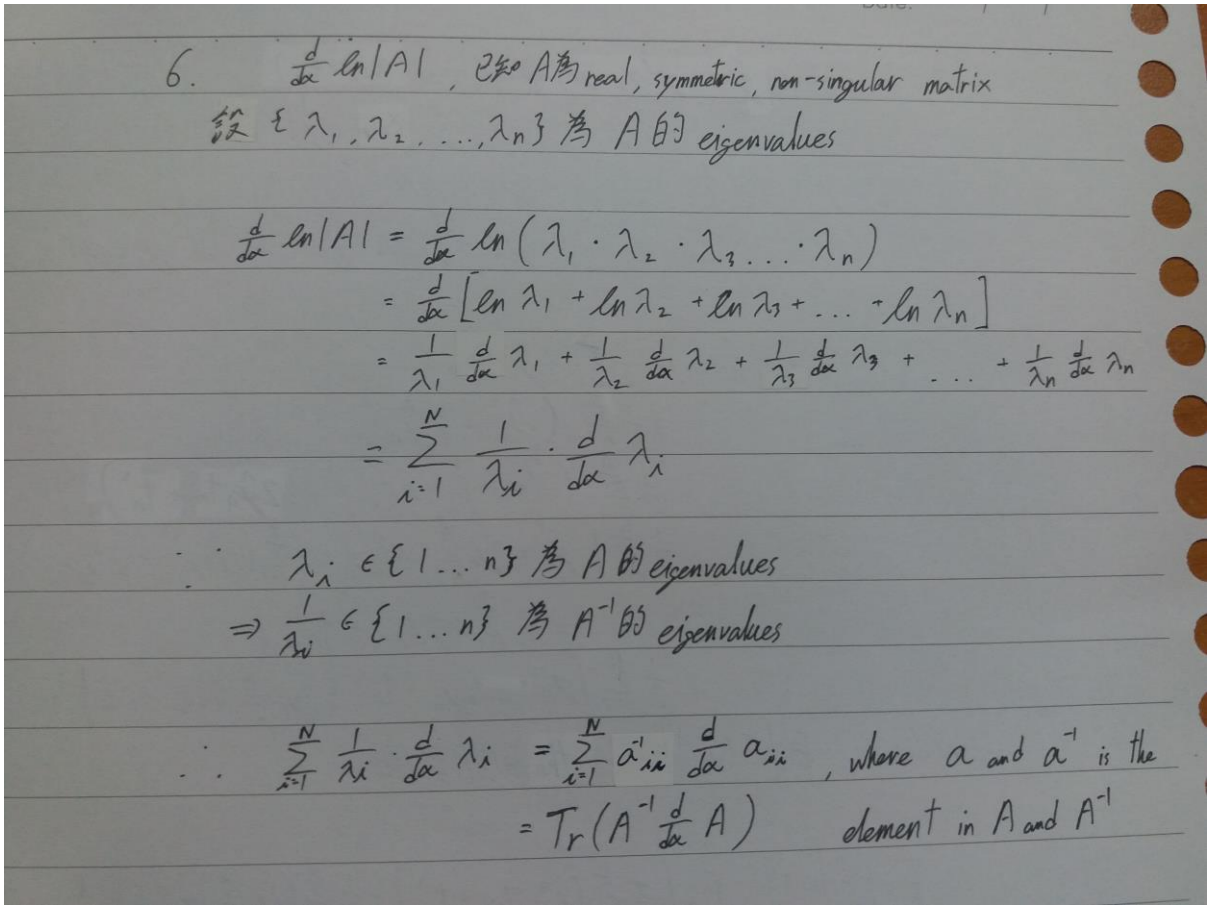
$\therefore E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left[2y_n \sum_{i=1}^{D} w_i \epsilon_i\right]\right] = 0$ and $E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left[-2t_n \sum_{i=1}^{D} w_i \epsilon_i\right]\right] = 0$

Also $E_x[\epsilon_i \epsilon_j] = \delta_{ij}\sigma^2$ 且 $\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

$$\Rightarrow E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2\right] = E_x\left[\frac{1}{2}\sum_{n=1}^{N}\left(\sum_{i=1}^{D}\sum_{j=1}^{D} w_i \epsilon_i w_j \epsilon_j\right)\right], \text{ where } i \text{ and } j \text{ 是獨立}$$
$$= \frac{1}{2}\sum_{i=1}^{D} w_i w_j \delta_{ij}\sigma^2 \quad , \text{where } i = j$$
$$= \frac{1}{2}\sum_{i=1}^{D} w_i^2 \sigma^2$$

$$\therefore E_x[\hat{E}(w)] = E_x[E(w)] + \frac{1}{2}\sum_{i=1}^{D} w^2 \sigma^2$$

6



6. $\frac{d}{d\alpha}\ln|A|$, 已知 A 為 real, symmetric, non-singular matrix

設 $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ 為 A 的 eigenvalues

$$\frac{d}{d\alpha}\ln|A| = \frac{d}{d\alpha}\ln(\lambda_1 \cdot \lambda_2 \cdot \lambda_3 \ldots \cdot \lambda_n)$$
$$= \frac{d}{d\alpha}[\ln\lambda_1 + \ln\lambda_2 + \ln\lambda_3 + \ldots + \ln\lambda_n]$$
$$= \frac{1}{\lambda_1}\frac{d}{d\alpha}\lambda_1 + \frac{1}{\lambda_2}\frac{d}{d\alpha}\lambda_2 + \frac{1}{\lambda_3}\frac{d}{d\alpha}\lambda_3 + \ldots + \frac{1}{\lambda_n}\frac{d}{d\alpha}\lambda_n$$
$$= \sum_{i=1}^{N}\frac{1}{\lambda_i}\cdot\frac{d}{d\alpha}\lambda_i$$

$\because \lambda_i \in \{1 \ldots n\}$ 為 A 的 eigenvalues

$\Rightarrow \frac{1}{\lambda_i} \in \{1 \ldots n\}$ 為 A^{-1} 的 eigenvalues

$\therefore \sum_{i=1}^{N}\frac{1}{\lambda_i}\cdot\frac{d}{d\alpha}\lambda_i = \sum_{i=1}^{N} a^{-1}_{ii}\frac{d}{d\alpha}a_{ii}$, where a and a^{-1} is the element in A and A^{-1}

$= Tr(A^{-1}\frac{d}{d\alpha}A)$