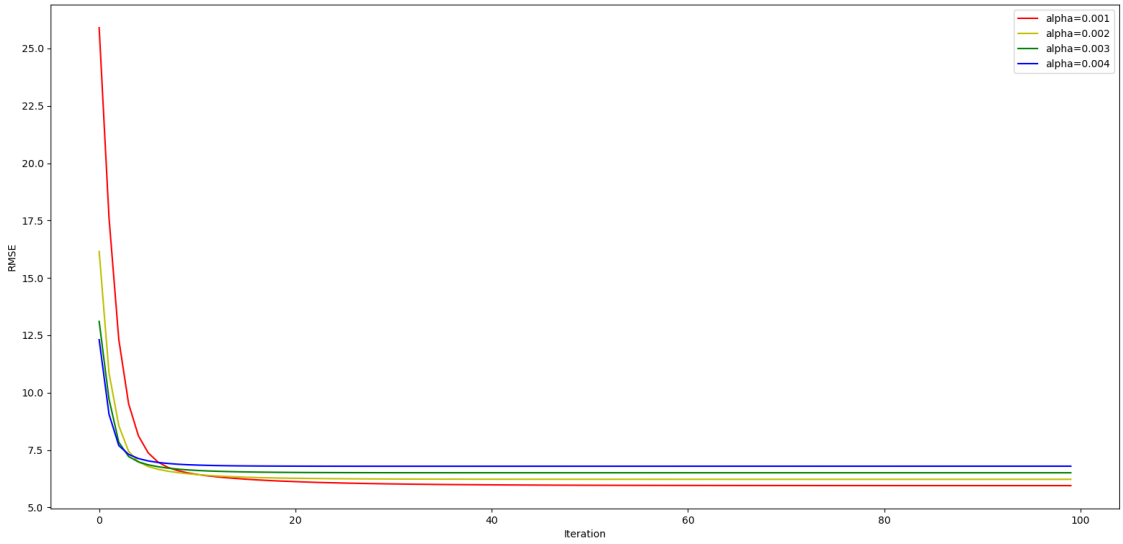
Homework 1 Report - PM2.5 Prediction

B05901022 電機三 許睿洋

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：



在 learning rate 較小的時候，初始的 loss 相當大，但其收斂到的數值卻是比較小的；反之，在 learnin rate 較大的時候，雖然有相當好的初始 loss，卻只能收斂到一個較大的數值。因此，慎選 learning rate 能夠確保訓練過程的 loss 能夠下降到較好的狀況。

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 Score	Private Score
With all features	7.65725	7.44538
Only with PM2.5	8.54986	8.37499

由於 PM2.5 的濃度還可能受到其他因素影響(例如:下雨)，只針對 PM2.5 進行訓練可能會有較差的結果。

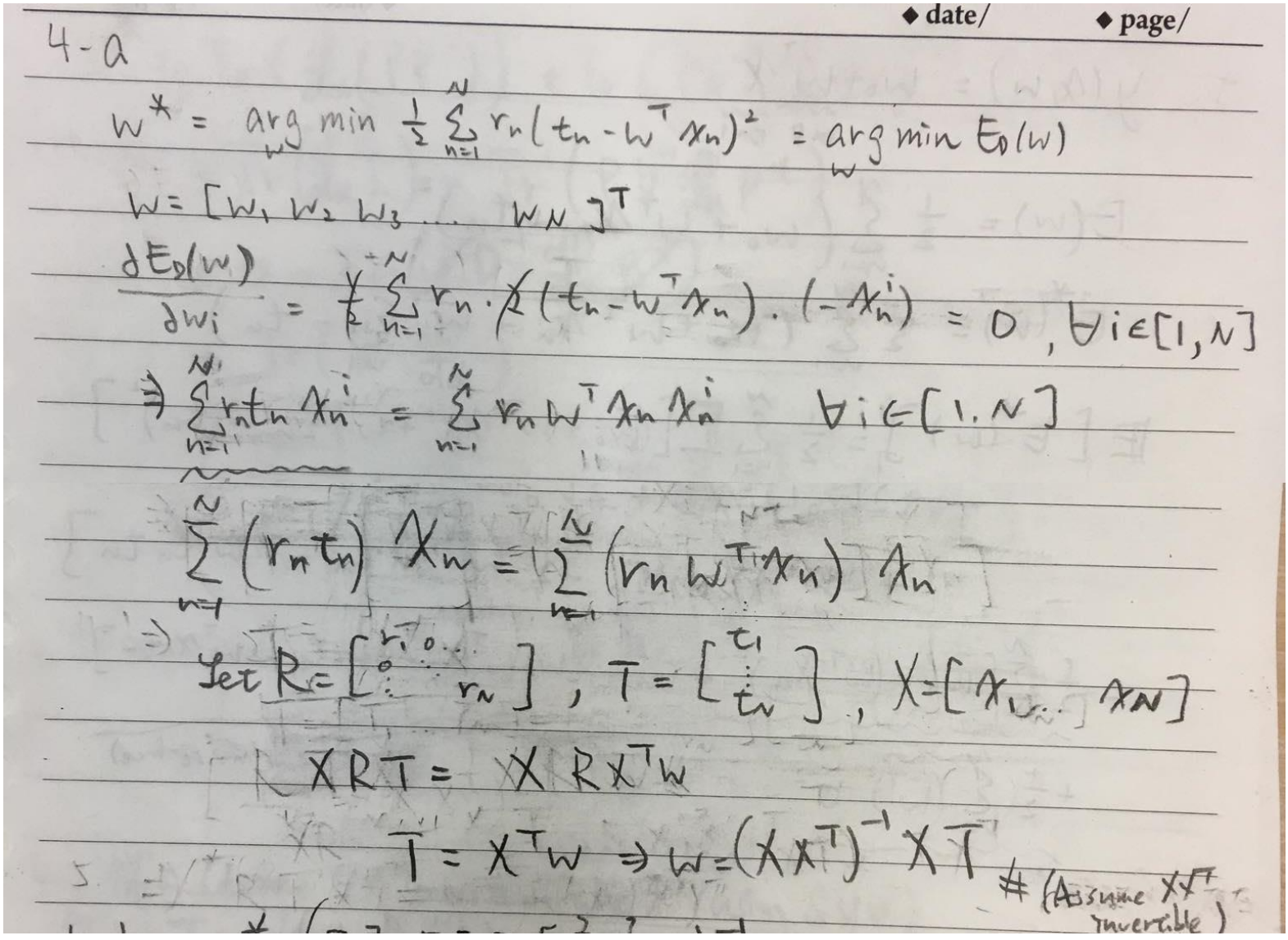
3. (1%)請分別使用至少四種不同數值的 regulization parameter λ 進行 training（其他參數需一至），討論及討論其 RMSE(traning, testing)（testing 根據 kaggle 上的 public/private
score）以及參數 weight 的 L2 norm。

	Training	Public	Private	L2Norm
$\lambda = 0.0001$	6.426675	7.66664	7.45865	15.669713
$\lambda = 0.001$	6.537951	7.71292	7.36896	15.019553
$\lambda = 0.01$	7.280320	7.97906	7.56481	12.144796
$\lambda = 0.1$	11.062287	9.96809	8.06805	8.504792

加入 regularization 可以使 loss 在訓練過程中能夠有相對穩定的變化，但是在這次的 case 中，加入過大的λ可能會使訓練結果變差，且可以從結果看出來可能是 L2Norm 被λ大幅壓縮所導致的惡化現象，因此如若要考慮加入 regularization，慎選參數是必要的。

4~6 數學題:

4.Collaborator: B05901009 高瑋聰、B05901034 劉奎元



4-2 補: (Assume XX^T non-invertible) (using SVD)

(a) $T = X^T w$

SVD $\Rightarrow$ $X^T = U \Sigma V^T$

$(X^T)^\dagger = V \Sigma^\dagger U^T$

$\Rightarrow w = (X^T)^\dagger T = V \Sigma^\dagger U^T T$ #

4-b $w^* = \left(\begin{bmatrix} 2 & 5 & 5 \\ 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 5 & 1 \\ 5 & 6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & 5 & 5 \\ 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}$

$= \begin{bmatrix} 54 & 41 \\ 41 & 46 \end{bmatrix}^{-1} \begin{bmatrix} 95 \\ 40 \end{bmatrix} = \begin{bmatrix} \frac{1910}{803} \\ \frac{-915}{803} \end{bmatrix}$

5.Collaborator:(Self only)

3. $y(x, w) = w_0 + \underbrace{w^T}_{1\times D} \underbrace{x}_{D\times 1}$

$E(w) = \frac{1}{2} \sum_{n=1}^{N} (w_0 + w^T x_n - t_n)^2$

$E^*(w) = \frac{1}{2} \sum_{n=1}^{N} (w_0 + w^T x_n + w^T \underbrace{\varepsilon_n}_{D\times 1} - t_n)^2$

$\mathbb{E}[E^*(w)] = \frac{1}{2} \sum_{n=1}^{N} \mathbb{E}[(w_0 + w^T x_n + w^T \varepsilon_n - t_n)^2]$

$= \frac{1}{2} \sum_{n=1}^{N} \mathbb{E}[w_0^2 + (w^T x_n)^2 + (w^T \varepsilon_n)^2 + t_n^2$

$+ 2 w_0 w^T x_n + 2 w_0 w^T \varepsilon_n - 2 w_0 t_n$

$+ 2 w^T x_n w^T \varepsilon_n - 2 w^T x_n t_n - 2 w^T \varepsilon_n t_n]$

$= \frac{1}{2} \sum_{n=1}^{N} [w_0^2 + (w^T x_n)^2 + t_n^2 + 2 w_0 w^T x_n - 2 w_0 t_n - 2 w^T x_n t_n]$ (original (noise-free))

$+ \frac{1}{2} \sum_{n=1}^{N} (w^T)^2 \sigma^2$ (noise regularization)

6.Collaborator:
B05901009 高瑋聰、B05901092 歐瀚墨、B05901011 許秉倫、
B05901034 劉奎元、B05901082 楊晟甫、B05901101 陳泓廷