ML HW1 Report

學號：R08942025　系級：　電信碩一　 姓名：徐瑋辰

請實做以下兩種不同 feature 的模型，回答第 1 ~ 2 題：

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

1. (1%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

batch	iteration	(1)	private	public	(2)	private	public
64	1000		5.57032	5.69377		109.23186	108.81695
64	500		5.52309	5.82427		54.81392	54.5677
64	100		5.86211	6.11524		12.56223	12.39876
32	1000		5.87311	6.34124		211.06645	210.51651
32	500		5.56471	5.74958		104.96767	104.66424
32	100		5.66771	5.99387		21.74261	21.51316
128	1000		5.57947	5.76018		55.64129	55.35397
128	500		5.4887	5.60404		28.52855	28.3078
128	100		5.70642	5.9161		8.60207	8.48891
128	50		5.47851	5.60307		6.86523	6.84066

觀察：

a. 在同樣的 batch 和 iteration 下，(1)的誤差相較於(2)較小

b. 當 batch 提高、iteration 減少時，(2)可有效降低誤差，但(1)則不明顯(可能已達到平衡)

c. 由此可知，其他汙染源之數據與 PM2.5 間，應考慮其 covariance

2. (1%)解釋什麼樣的 data preprocessing 可以 improve 你的 training/testing accuracy，ex. 你怎麼挑掉你覺得不適合的 data points。請提供數據(RMSE)以佐證你的想法。

Idea:從 PM2.5 數值中，刪除數值過大與過小者

Criteria: $1 < y < 80$

Result:

128	1000		5.60404	5.78372		55.61288	55.34186

Conclusion: 與上一題相比，沒有太多進步的空間

3.(3%) Refer to math problem

自下頁起

1. (a)

$$L_{ssq}(w,b) = \frac{1}{2\times 5}\sum_{i=1}^{5}(y_i-(w x_i+b))^2$$

$$\frac{\partial L}{\partial w} = \sum_{i=1}^{5} 2(y_i-(wx_i+b))(-x_i) = 0$$

$$\Rightarrow (1.2-(w+b)) + 2(2.4-(2w+b)) + 3(3.5-(3w+b)) + 4(4.1-(4w+b)) + 5(5.6-(5w+b)) = 0$$

$$\Rightarrow 60.9 - 55w - 15b = 0 \quad \text{— ①}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{5} 2(y_i-(wx_i+b)) = 0$$

$$\Rightarrow (1.2-(w+b)) + (2.4-(2w+b)) + (3.5-(3w+b)) + (4.1-(4w+b)) + (5.6-(5w+b)) = 0$$

$$\Rightarrow 16.8 - 15w - 5b = 0 \quad \text{— ②}$$

by ① & ② $\Rightarrow$ $\underline{w = 1.05 \quad b = 0.21}$ #

1 (b)

$$L_{ssq}(\underline{w},b) = \frac{1}{2N}\sum_{i=1}^{N}(y_i-(w^T x_i+b))^2$$

Let $b = w_0 x_0 \Rightarrow \tilde{w} = \begin{bmatrix} b \\ w \end{bmatrix}$

$$\Rightarrow L(\tilde{w}) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - x_i^T\tilde{w})^2$$

$$\frac{\partial L}{\partial \tilde{w}} = \frac{1}{N}\sum_{i=1}^{N}(x_i^T\tilde{w} - y_i)x_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i^T\tilde{w}x_i = \sum_{i=1}^{N} y_i x_i$$

$$\Rightarrow \tilde{w} = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^T x_i} \Rightarrow \underline{(X^TX)^{-1}X^Ty}$$ #

1(c)

$$L_{reg}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - (w^T x_i + b)\right)^2 + \frac{\lambda}{2}\|w\|^2$$

let $\tilde{w} = \begin{bmatrix} b \\ w \end{bmatrix} \in R^{N+1}$

$$\Rightarrow L(\tilde{w}) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \tilde{w}^T x_i)^2 + \frac{\lambda}{2}\|\tilde{w}\|^2 - b^2$$

$$\frac{\partial L}{\partial \tilde{w}} = \frac{1}{\cancel{2}N}(-\cancel{2})\sum_{i=1}^{N}(y_i - \tilde{w}^T x_i)x_i + \lambda\tilde{w} = 0$$

$$\frac{1}{N}(-X^T y + X^T X\tilde{w}) + \lambda\tilde{w} = 0$$

$$\Rightarrow X^T X\tilde{w} + \lambda N\tilde{w} = X^T y$$

$$\Rightarrow \tilde{w}(X^T X + \lambda N I^N) = X^T y$$

$$\Rightarrow \underline{\tilde{w} = (X^T X + \lambda N I)^{-1} X^T y} \ \#$$

(I 為單位矩陣)

2.

$$L_{ssq}(w,b) = \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i + \eta_i) - y_i\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i) + f_{w,b}(\eta_i) - y_i\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(\underbrace{f_{w,b}(x_i) - y_i}_{A} + \underbrace{f_{w,b}(\eta_i)}_{B}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i) - y_i\right)^2 + f_{w,b}(\eta_i)^2 + 2AB\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i) - y_i\right)^2\right] + \underline{\mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N} f_{w,b}(\eta_i)^2\right] + \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N} 2AB\right]}$$

$$\Rightarrow \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N} f_{w,b}(\eta_i)^2\right] + \cancel{\mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N} 2\left(f_{w,b}(x_i) - y_i\right)\underbrace{f_{w,b}(\eta_i)}_{=0}\right]}$$

$$\Rightarrow \mathbb{E}\left[\frac{1}{2N}\cdot N\cdot(w^T\eta_i + b)^2\right]$$

$$\Rightarrow \mathbb{E}\left[\frac{1}{2}(\eta w + b)^T(\eta w + b)\right]$$

$$\Rightarrow \frac{1}{2}\mathbb{E}\left[\|\eta w\|^2\right]$$

$$\Rightarrow \frac{1}{2}\mathbb{E}[\eta]\|w\|^2$$

$$\Rightarrow \underline{\frac{\sigma^2}{2}\|w\|^2} \ \#$$

3. (a)

$$e_k = \frac{1}{N}\sum_{i=1}^{N}\left(g_k(x_i) - y_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(g_k^2(x_i) + y_i^2 - \underline{2g_k(x_i)y_i}\right)$$

$$\sum_{i=1}^{N} g_k(x_i)y_i = \frac{1}{2}\left(\sum_{i=1}^{N} g_k^2(x_i) + \sum_{i=1}^{N} y_i^2 - Ne_k\right)$$

$$= \frac{1}{2}\left(Ns_k + Ne_0 - Ne_k\right)$$

$$= \frac{N}{2}\left(s_k + e_0 - e_k\right)$$

#

3 (b)

$$L = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right)^2$$

$$\frac{\partial L}{\partial \alpha_k} = \frac{2}{N}\sum_{i=1}^{N} g_k(x_i)\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right) = 0$$

$$\Rightarrow \cancel{\frac{2}{N}}\sum_{k=1}^{K}\underbrace{\left(\sum_{i=1}^{N} g_k^2(x_i)\right)}_{Ns_k}\alpha_k - \cancel{\frac{2}{N}}\underbrace{\sum_{i=1}^{N} g_k(x_i)y_i}_{\frac{N}{2}(s_k+e_0-e_k)\ (\text{by } 3(a))} = 0$$

$$\therefore\ \alpha_k \propto \frac{s_k+e_0-e_k}{2s_k}$$

#