學號：B05705006 系級： 資管四 姓名：李和維

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)

備註：

a. NR請皆設為0，其他的非數值(特殊字元)可以自己判斷

b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

c. 第1-2題請都以題目給訂的兩種model來回答

d. 同學可以先把model訓練好，kaggle死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

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

	抽全部9小時內的污染源	抽全部9小時內pm2.5
RMSE Public	5.52094	5.95715
RMSE Private	5.38504	5.90364

根據上表，可以得知抽取全部 9 小時內的污染源的一次項並加上 bias 做為 linear regression model 的 features，無論是 public 或 private，RMSE 比起只抽取 pm2.5 加上 bias 的模型都有顯著下降。

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。

- 對於丟失值的處理嘗試以下實驗，其中 iteration 次數皆設為 5000
    - 將 nan 或空白值改成 0

	抽全部9小時內的污染源	抽全部9小時內pm2.5
RMSE Training	4.106373	4.384693
RMSE Public	5.57511	5.95670
RMSE Private	5.38319	5.90810

-
    - 將 nan 或空白值改成前一格之值

	抽全部9小時內的污染源	抽全部9小時內pm2.5
RMSE Training	3.920856	4.370559
RMSE Public	5.90681	5.95715

RMSE Private	5.74099	5.90364

- 對於 outlier 的處理嘗試以下幾種實驗，其中 iteration 次數也皆設為 5000，outlier 定義為：距離該資料種類平均超過兩倍標準差。
    - 移除含有 outliers 的 training data

	抽全部9小時內的污染源	抽全部9小時內pm2.5
RMSE Training	4.076883	4.384693
RMSE Public	5.63991	5.95670
RMSE Private	5.41116	5.90810

-
    - 將 outliers 置換為平均

	抽全部9小時內的污染源	抽全部9小時內pm2.5
RMSE Training	4.106373	4.619356
RMSE Public	5.57511	5.93199
RMSE Private	5.38319	5.84156

由以上實驗可以發現，對此筆 dataset 而言若把丟失值換成上一個小時的值，雖然 training score 會最低但 testing 時都會比較高，有 overfitting 的問題。所以最終選擇使用補 0 的方式效果較佳。
另外，處理 outlier 的方式中，把 outlier 置換為平均值出來的結果明顯比直接移除該 training data 要好很多，推斷是因為 training data 數量要多才能 train 出較精確的 model 。

3.(3%) Refer to math problem

No.

Date

1-(a)

$$w = \frac{\sum (x_n - \bar{x})(y_n - \bar{y})}{\sum (x_n - \bar{x})^2} = \frac{2\times 2.16 + 1\times 0.96 + 0 + 1\times 0.74 + 2\times 2.24}{4+1+0+1+4}$$

$$= \underline{\underline{1.05}}, \quad b = \bar{y} - w\times\bar{x} = 3.36 - 1.05\times 3 = \underline{\underline{0.21}}$$

$$\underline{(w, b) = (1.05, 0.21)} \quad ※$$

1-(b)

$$\frac{\partial L}{\partial b} = \frac{-1}{N}\sum_{n=1}^{N}(y_n - (w^T x_n + b)) = 0$$

$$Nb = \sum_{n=1}^{N} y_n - \sum_{n=1}^{N} w^T x_n, \quad \underline{b = \bar{y} - w^T\bar{x}}$$

$$\frac{\partial L}{\partial w} = \frac{-1}{N}\sum_{n=1}^{N}(y_n - (w^T x_n + b))\, x_n = 0,$$

$$\sum_{n=1}^{N}(y_n - w^T x_n - \bar{y} + w^T\bar{x})\, x_n = 0,$$

$$\sum y_n x_n - \sum \bar{y} x_n + \sum w^T \bar{x} x_n - \sum w^T x_n^2 = 0,$$

$$w^T = \frac{\bar{y}\sum x_n - \sum x_n y_n}{\bar{x}\sum x_n - \sum x_n^2} = \frac{n\bar{y}\bar{x} - \sum x_n y_n}{n\bar{x}^2 - \sum x_n^2} = \underline{\underline{\frac{\sum_{n=1}^{n}(x_n - \bar{x})(y_n - \bar{y})}{\sum_{n=1}^{n}(x_n - \bar{x})^2}}}$$

$$\underline{(w, b) = \left(\frac{\sum_{n=1}^{n}(x_n - \bar{x})(y_n - \bar{y})}{\sum_{n=1}^{n}(x_n - \bar{x})^2},\ \bar{y} - w^T\bar{x}\right)} \quad ※$$

1-(c)

$$\frac{\partial L}{\partial b} = 0 \Rightarrow b = \bar{y} - w^T \bar{x} ,$$

$$\frac{\partial L}{\partial w} = \frac{-1}{N} \Sigma (y_n - w^T x_n - \bar{y} + w^T \bar{x}) x_n + \lambda w = 0 ,$$

$$\Sigma w^T \bar{x} x_n - \Sigma w^T x_n^2 - N\lambda w = \Sigma \bar{y} x_n - \Sigma y_n x_n$$

$$w = \frac{\Sigma (\bar{y} - y_n) x_n}{\Sigma (\bar{x} - x_n) x_n - \lambda N} = \frac{\sum_{n=1}^{N} (y_n - \bar{y})(x_n - \bar{x})}{\sum_{n=1}^{N} (x_n - \bar{x})^2 + \lambda N}$$

2.

$$E\left[\frac{1}{2N} \sum_{n=1}^{N} (f(x_n + \eta_n) - y_n)^2 \right]$$

$$= E\left[\frac{1}{2N} \sum_{n=1}^{N} (w^T x_n + w^T \eta_n + b - y_n)^2 \right]$$

$$= E\left[\frac{1}{2N} \Sigma (f(x_n) + w^T \eta_n - y_n)^2 \right]$$

$$= E\left[\frac{1}{2N} \left(\Sigma (f(x_n) - y_n)^2 + 2 \Sigma (f(x_n) - y_n)(w^T \eta_n) + \Sigma (w^T \eta_n)^2 \right) \right]$$

$$= \frac{1}{2N} \left(\Sigma (f(x_n) - y_n)^2 + \|w\|^2 N \sigma^2 \right.$$

$$= \frac{1}{2N} \sum_{n=1}^{N} (f_{w,b}(x_n) - y_n)^2 + \frac{\sigma^2}{2} \|w\|^2$$

3-(a)

$$N e_k = \Sigma (g_k(x_n) - y_n)^2 = \Sigma g_k(x_n)^2 - 2 \Sigma g_k(x_n) y_n + \Sigma y_n^2$$

$$= N s_k - 2 \Sigma g_k(x_n) y_n + N e_0 ,$$

$$\Rightarrow \sum_{n=1}^{N} g_k(x_n) y_n = \frac{N(s_k + e_0 - e_k)}{2}$$

(b)

$$\min\left[\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_n)-y_n\right)^2\right]$$

$$=\min\left[\frac{1}{N}\sum_{n=1}^{N}\left(\left(\sum_{k=1}^{K}\alpha_k g_k(x_n)\right)^2-2\cdot\sum_{k=1}^{K}\alpha_k g_k(x_n)\cdot y_n+y_n^2\right)\right]$$

$$=\min\left[\frac{1}{N}\left(\sum_{n=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_n)\right)^2-2\cdot\sum_{k=1}^{K}\sum_{n=1}^{N}\alpha_k g_k(x_n)y_n+\sum_{n=1}^{N}y_n^2\right)\right]$$

$$=\min\frac{1}{N}\left(\sum\left(\sum\alpha_k g_k(x_n)\right)^2-\cancel{2}\cdot\sum_{k=1}^{K}\alpha_k\cdot\left(\frac{N(s_k+e_0-e_k)}{\cancel{2}}\right)+\sum_{n=1}^{N}y_n^2\right)$$

$\Rightarrow$ 因 $g_k(x_n)$, N, s_k, e_0, e_k, $\sum_{n=1}^{N}y_n^2$ 皆為已知(視為常數)

原式 = minimize 一個 α_k 二次多項式，可 obtain the optimal weights $\alpha_1 \ldots \alpha_K$ ※