學號：R06943148 系級： 電子碩一 姓名：謝孟洋

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

(1) 抽全部 9 小時內的污染源 feature 的一次項(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，其他的數值不要做任何更動

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

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

a.(所有 feature) $\sqrt{(7.46237^2 + 5.53562^2) \div 2} = 6.57001$

b.(只取 pm2.5) $\sqrt{(7.44013^2 + 5.62719^2) \div 2} = 6.59624$

討論：雖然只取 pm2.5 的結果變差了，不過其實變化非常微小，可能可以視為其實這兩種取 feature 的方式不對結果造成太大的影響。

2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

a. $\sqrt{(7.65925^2 + 5.44092^2) \div 2} = 6.64333$

b. $\sqrt{(7.57904^2 + 5.79187^2) \div 2} = 6.74491$

討論：可以見到改成只抽前 5 小時的結果，相較原本抽 9 小時(第一題結果)，分數都變差了，所以應該還是得要取 9 小時。原因可能是因為如果較久以前的資料比較不重要，在 training 的過程中 weight 自然會降下來。

3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

4. (1%)在線性回歸問題中，假設有 N 筆訓練資料，每筆訓練資料的特徵 (feature) 為一向量 x^n，其標註(label)為一存量 y^n，模型參數為一向量 w (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為$\sum_{n=1}^{N} (y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \dots y^N]^T$ 表示，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請寫下算式並選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$

(b) $(X^TX)^{-0}X^Ty$

(c) $(X^TX)^{-1}X^Ty$

(d) $(X^TX)^{-2}X^Ty$