學號：R06922002 系級： 資工碩一 姓名：姚嘉昇

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

	Private Score	Public Score	Average Score
18 feature, 9 hours	5.63635	8.01539	6.825870
1 feature, 9 hours	5.83648	7.39590	6.616190

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

	Private Score	Public Score	Average Score
18 feature, 5 hours	5.40360	7.71723	6.560415
1 feature, 5 hours	5.86272	7.52172	6.692220

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

	λ	Private Score	Public Score	Average Score
18 feature, 9 hours	0.1	5.63627	8.01304	6.824655
	0.01	5.63637	8.01492	6.825645
	0.001	5.63635	8.01532	6.825835
	0.0001	5.63635	8.01536	6.825855
1 feature, 9 hours	0.1	5.84309	7.40644	6.624765
	0.01	5.83714	7.39696	6.617050
	0.001	5.83655	7.39601	6.616280
	0.0001	5.83649	7.39591	6.616200

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$\sum_{n=1}^{N} (y^n - x^n \cdot w)^2$。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) $(X^TX)^{-2}X^Ty$

答案：(c)

證明 1：

證明 2：

為了使損失函數最小化，必須讓 $y - Xw = 0$，則 $y = Xw$

兩邊同乘 X^T 後，則 $X^Ty = X^TXw$

兩邊同乘$(X^TX)^{-1}$ 後，則 $(X^TX)^{-1}X^Ty = (X^TX)^{-1}X^TXw$

右式整理後，可得 $w = (X^TX)^{-1}X^Ty$