學號：R06922129 系級： 資工碩一 姓名：丁縉楷

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	public	RMSE
所有feature	5.50413	7.83378	6.769917139
pm2.5	5.62719	7.44013	6.596241419

在取9小時的sample中，pm2.5的表現較好，推測在同樣的iteration次數下取全部的feature可能會有overfit的效果，所以取18個feature反而比只取pm2.5還要差

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

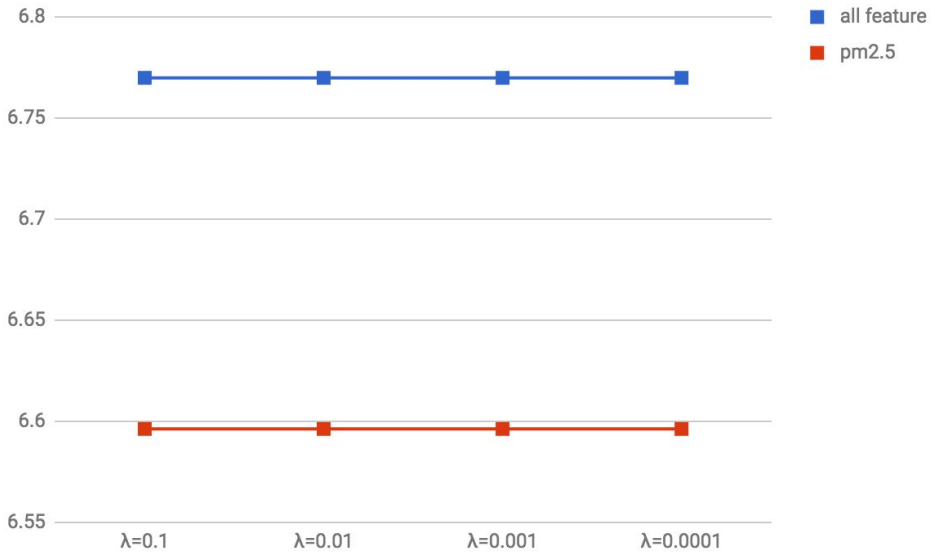
	private	public	RMSE
所有feature	5.37815	7.73754	6.663108234
pm2.5	5.79187	7.57904	6.744909392

取5個小時的話，所有feature的表現又比只取pm2.5還好，所以跟上面相反，這次pm2.5 feature取太少反而underfit，比較取9 or 5小時的差別，會發現取所有feature error是降低，只取pm2.5是升高的，所以驗證所有feature取9小時overfit，pm2.5取5小時underfit的推論。

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

	private	public	RMSE
all feature(λ=0.1)	5.50414	7.83378	6.769921204
pm2.5(λ=0.1)	5.6272	7.44012	6.596240045
all feature(λ=0.01)	5.50413	7.83378	6.769917139
pm2.5(λ=0.01)	5.62719	7.44013	6.596241419
all feature(λ=0.001)	5.50413	7.83378	6.769917139

pm2.5(λ=0.001)	5.62719	7.44013	6.596241419
all feature(λ=0.0001)	5.50413	7.83378	6.769917139
pm2.5(λ=0.0001)	5.62719	7.44013	6.596241419



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

Ans:(c)

loss function: $(wX-y)^t(wX-y)$

最小值發生在偏微分為0的地方：

$$\frac{\partial X}{\partial w} = 2X^TXw - 2X^Ty = 0$$

$$X^TXw = X^Ty$$

$$w = (X^TX)^{-1}X^Ty$$