學號：R07922134 系級： 資工碩一 姓名：陳紘豪

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

c. 第1-3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. (2%)記錄誤差值 (RMSE)(根據kaggle public+private分數)，討論兩種feature的影響

此為採用 lr=0.01, epoch=20000, optimizer=Adam 的結果

	Public	Private
(1) All features	5.81763	7.28175
(2) Only PM2.5	5.93022	7.24763

討論：

在 Public 上的 RMSE 是 (1) 表現的比 (2) 稍微好一些些，但是在 Private 上 (2) 贏過 (1)，這樣的原因可能是因為採用 All features 來訓練模型的話可能考慮了太多不必要的因素，以至於 Model 在預測 PM2.5 時產生較大的誤差。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

All features		
	Public	Private
(1) 9 hours	5.81763	7.28175
(2) 5 hours	6.00889	7.24587

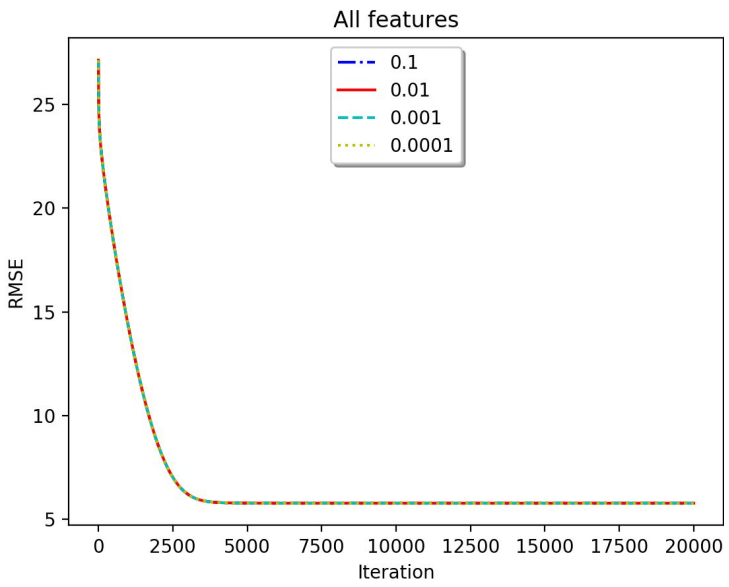
Only PM2.5		
	Public	Private
(1) 9 hours	5.93022	7.24763
(2) 5 hours	6.23692	7.24509

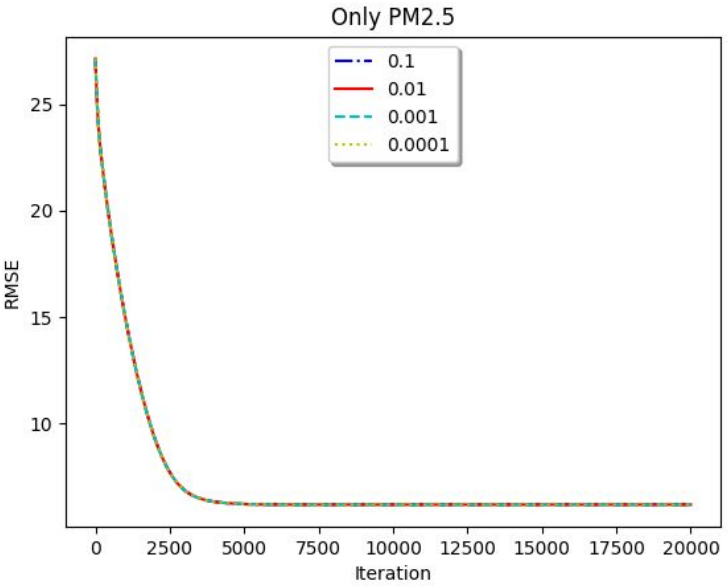
討論：

在 Public 上，不管取的 features 是什麼，採用前 5 hrs 的表現會比採用前 9hrs 來的更差，可能是因為資料量太少的關係，所以導致 training 出來的結果不盡理想，面臨到了 underfitting 的狀況。

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

此為採用 lr=0.01, epoch=20000, optimizer=Adam 的結果





討論：

在兩張圖當中，不同的 λ 中 RMSE 隨著 iteration 降低的曲線幾乎都是一模一樣的，可以想像成是因為，不管是採用全部 features 或是只有 pm2.5，我們的 function set 裏頭都是一次式所組成的，所以對於 regularization 所想要找的「較平滑」的線這個目標來說，對於實際上效能的影響就沒有那麼大，因為 function set 本身就早都是平滑的，因此如果 fuction set 有包含中 2 次以上的 function（例如 features 中有用到 pm2.5 的二次項），regularization 可能就會比較有實際影響。

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)為一純量 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 ？請選出正確答案。(其中X^TX為invertible)

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

(b) $(X^TX)yX^T$

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

(d) $(X^TX)^{-1}yX^T$

答案為(c)