學號：B04901102 系級： 電機三 姓名：簡仲明

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

全部汙染源：7.90808+5.55183=13.45991

僅抽PM2.5：7.92794+5.86876=13.79670

抽取全部汙染源較為準確，但兩者相差不如想像中大。推測原因是因為全部汙染源中雜有許多跟PM2.5無關的數據，且某些數據含有invalid的值(ex. PM2.5=-1)。若刪去一些無關的項，並對invalid的值做適當處理後，最佳可以達到大約7.28454+5.78131=13.06585之譜。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

全部feature：7.73453+5.62488=13.35941

僅抽PM2.5：8.15074+6.06706=14.21780

整體而言取5小時的準確度較取9小時差，特別是只抽PM2.5的情況下，因為只剩下5個維度，因此準確度下降更明顯。

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

λ = 0.1

全部汙染源：7.89936+5.50076=13.40012

僅抽PM2.5：7.89712+5.83131=13.72923

λ = 0.01

全部汙染源：7.80826+5.57808=13.38634

僅抽PM2.5：7.86822+5.79722=13.66544

λ = 0.001

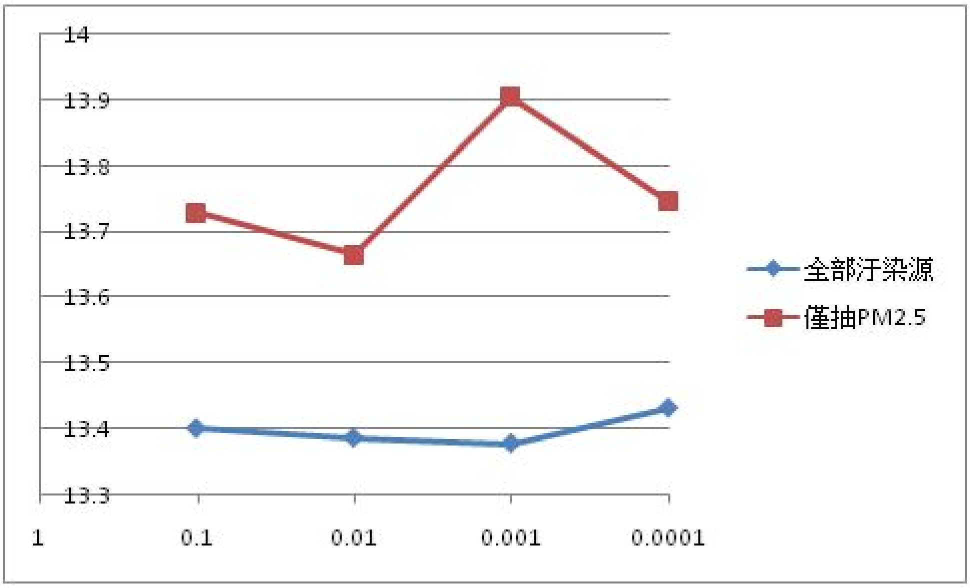
全部汙染源：7.80787+5.57007=13.37794

僅抽PM2.5：7.98257+5.92065=13.90322

λ = 0.0001

全部汙染源：7.93013+5.50092=13.43105

僅抽PM2.5：7.90312+5.84268=13.74580



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 $X = [x^1\ x^2\ \ldots\ x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2\ \ldots\ y^N]^T$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) $(X^TX)^{-2}X^Ty$

正解為(c)。

預測值可表示為一向量$w1x^1+w2x^2$......此向量可以理解為所有X中的column vector的某一線性組合，亦即線性空間W=Span($\{x^1,x^2......\}$)中的某一向量。接著注意loss function的定義，即為y 向量與預測向量的Eucliid distance的平方。欲使得loss最小，則我們必須找到W中與y距離最近的向量即為y在W上的投影。

接著帶入projection matrix的公式$P=X(X^TX)^{-1}X^T$，我們可以得到 $Xw=Py=X(X^TX)^{-1}X^Ty$的結果，接著兩邊同時乘上$(X^TX)^{-1}X^T$，即可得到最終答案 $w=(X^TX)^{-1}X^Ty$。