學號：R07943107 系級： 電子所碩一 姓名：徐晨皓

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**

答:

|  |  |  |
| --- | --- | --- |
|  | Public score | Private Score |
| 所有 features | 5.63779 | 7.21546 |
| 只有 PM2.5 | 5.90263 | 7.22356 |

由實驗結果可知，只有PM2.5的預測誤差值較所有features誤差值大。這有可能是影響當前的PM2.5數值不僅只有先前的PM2.5數值，也包含其他元素的影響。由維基百科的資料顯示，PM2.5的成分可能由[硫](https://zh.wikipedia.org/wiki/%E7%A1%AB)和[氮](https://zh.wikipedia.org/wiki/%E6%B0%AE)的[氧化物](https://zh.wikipedia.org/wiki/%E6%B0%A7%E5%8C%96%E7%89%A9)轉化而成，因此空氣中[硫](https://zh.wikipedia.org/wiki/%E7%A1%AB)和[氮](https://zh.wikipedia.org/wiki/%E6%B0%AE)的氧化物濃度可能與PM2.5數值有關。

(備註: 所有實驗皆使用Adagrad; 初始學習率1.0; Iterations: 10000)

(資料來源: https://zh.wikipedia.org/wiki/%E6%87%B8%E6%B5%AE%E7%B2%92%E5%AD%90)

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

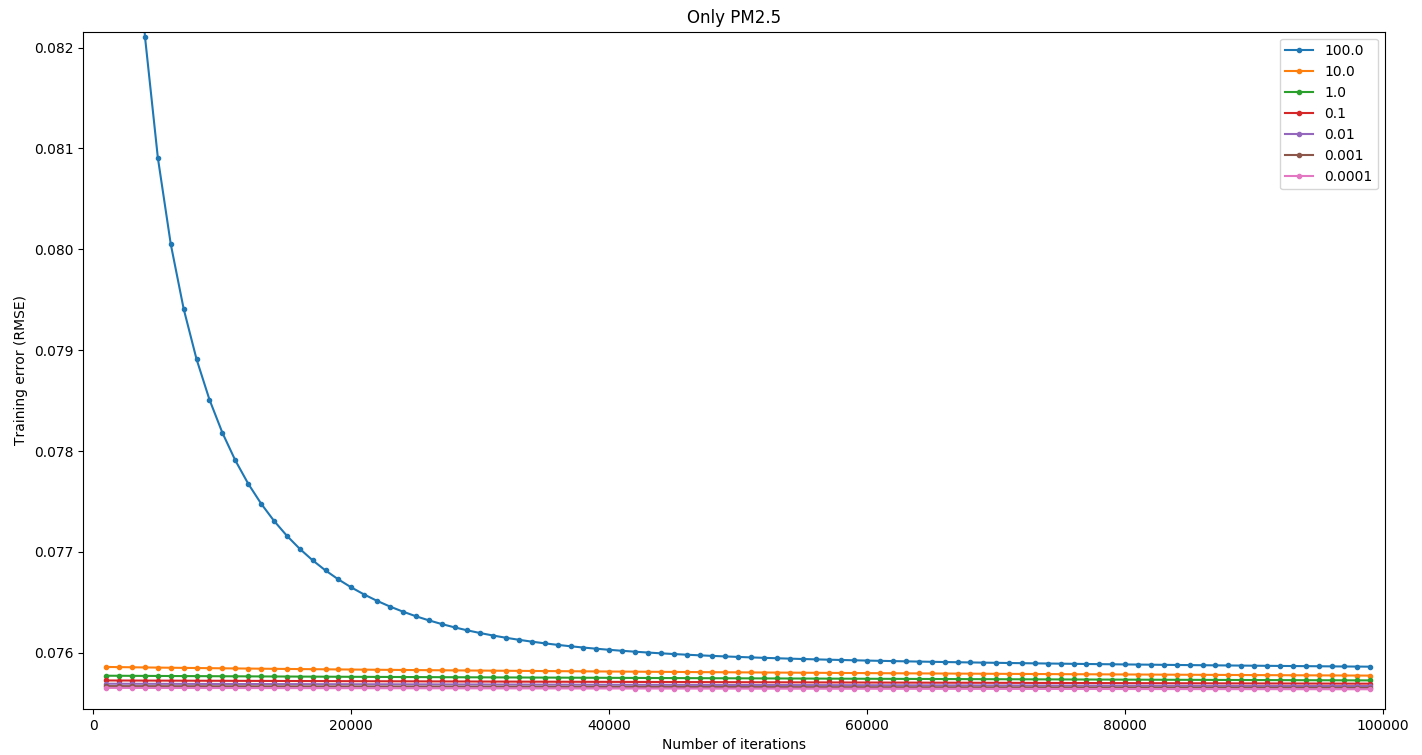
答:

|  |  |  |
| --- | --- | --- |
|  | Public score | Private Score |
| 所有 features + 抽前9小時 | 5.63779 | 7.21546 |
| 所有 features + 抽前5小時 | 5.98257 | 7.16701 |
| 只有 PM2.5 + 抽前9小時 | 5.90263 | 7.22356 |
| 只有 PM2.5 + 抽前5小時 | 6.22732 | 7.22552 |

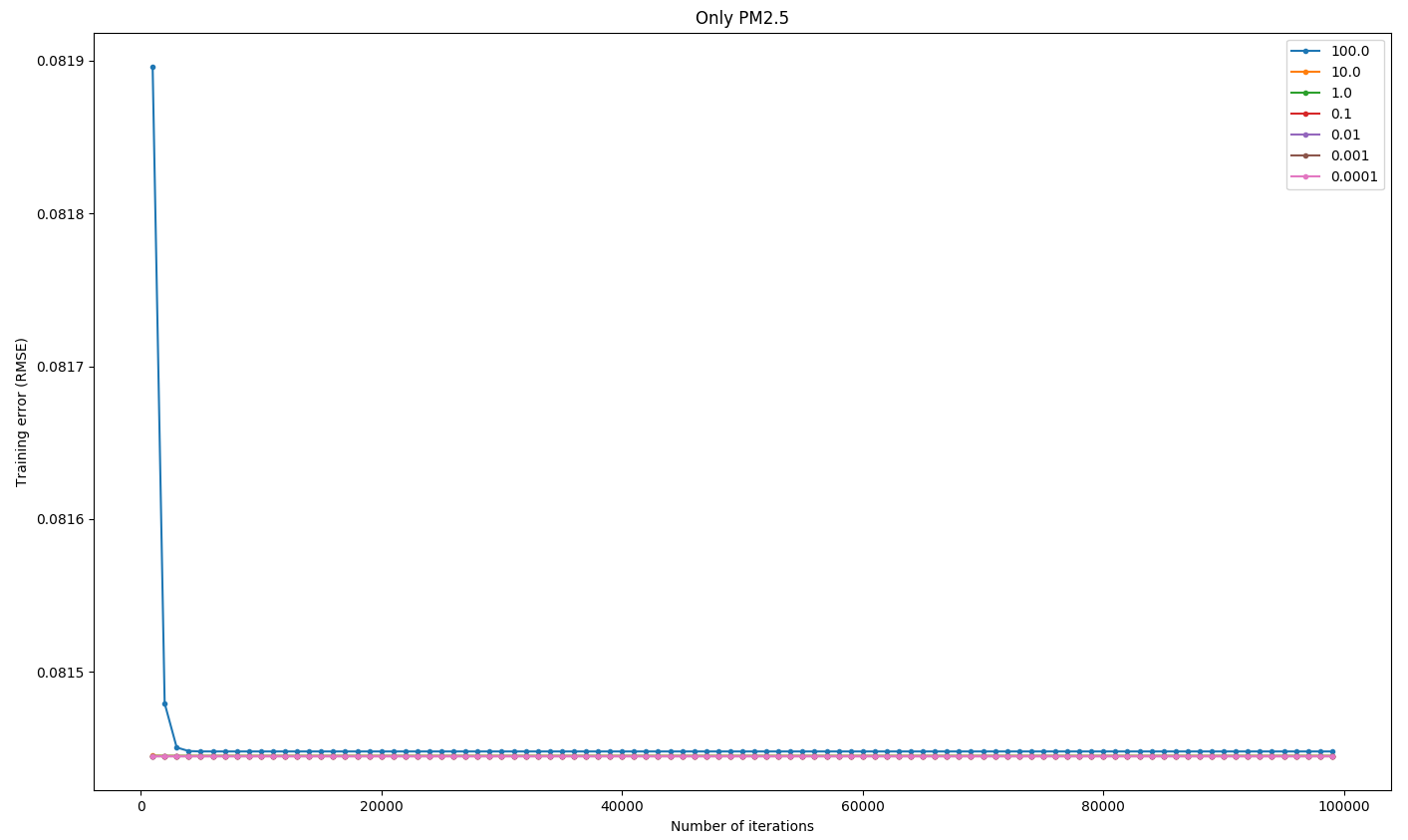
大致而言，抽取前9小時的誤差在public與private set上的誤差都比抽取前5小時的誤差小 (除了所有 features在private set上的表現)。以這結果而言，當前PM2.5的值與先前的PM2.5有一定的相關性，因此參考較多小時前的值，有助於降低預測誤差值。另外，在訓練過程中，由於抽取前5小時的features較少，因此training的時間有顯著的降低。

(備註: 所有實驗皆使用Adagrad; 初始學習率1.0; Iterations: 10000)

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

答:

**All features**



**Only PM2.5**

答: 由上二圖可知，當λ越大時，training error越大，隨著λ漸小，training error也隨之減小。我們知道regularization能使model對於input的變化減小，從而避免overfitting的狀況。因此當λ大時，model並不會趨向符合training data，這也就是為何λ越大，training error越大。另外，在all features的實驗中，不同的λ對應的error有顯著的不同；但在only PM2.5的實驗中，不同的λ對應的error並**沒有**顯著的不同。這可能是因為only PM2.5的model已經非常簡單，不像all features的model那樣複雜，所以不同的λ對於only PM2.5的實驗並沒有太大的影響。

(備註: 所有實驗皆使用Adagrad; 初始學習率1.0; Iterations: 10000)

**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 xn，其標註(label)為一純量 yn，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 。若將所有訓練資料的特徵值以矩陣 X = [x1 x2 … xN]T 表示，所有訓練資料的標註以向量 y = [y1 y2 … yN]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XTX為invertible)**

1. **(XTX)XTy**
2. **(XTX)yXT**
3. **(XTX)-1XTy**
4. **(XTX)-1yXT**

答: (c)