學號：R05922156 系級： 資工碩二 姓名：黃子瑋

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**

|  |  |  |  |
| --- | --- | --- | --- |
| feature | RMSE (kaggle) | RMSE (validation) | RMSE (training data) |
| 全部feature | 7.46740 | 5.056002 | 5.698215 |
| PM2.5 | 7.10458 | 5.537820 | 6.135208 |

討論:

因為10/12實還看不到kaggle的private分數，所以這邊我只使用了kaggle public的數據，上表的數據都是以adagrad加上同樣的trainning次數去看，以這個紀錄數據來看，只使用PM2.5的testing data結果是比較好的，雖然說在validation跟trainning data方面看起來都是用所有feature可以做到比較低，不過我的validation data也只切出120筆資料，跟kaggle上的public一樣，但我覺得有可能是因為資料量不夠大造成取樣上的偏差，才會導致kaggle的public跟validation差這麼多，所以照理來說應該是要讀取全部的feature會比較好，但因為在kaggle上是指讀取PM2.5比較好，所以我後來有嘗試另一個方法是只讀取10個feature，但做出來的結果感覺也跟上述的差不多

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

(因為沒有kaggle次數了，且開放無線上傳kaggle的那天剛好有事情無法使用，所以我事先用validation測試並記錄)

|  |  |  |  |
| --- | --- | --- | --- |
|  | feature | RMSE (validation) | RMSE (training data) |
| 全部feature | 9小時 | 5.056002 | 5.698215 |
| 5小時 | 5.509665 | 5.815375 |
| Only PM2.5 | 9小時 | 5.537820 | 6.135208 |
| 5小時 | 5.012593 | 6.230248 |

討論:

1. 全部feature

觀察到不管是在validation data還是在trainning data ，抽取9個小時的RMSE都比抽取5個小時的要好，跟預期中的差不多，trainning data越多維通常會得到較好的結果

1. Only PM2.5

但在PM2.5的validation中卻得出了相反的結果，只抽取5小時的validation 的RMSE反而比抽9小時的還要好，這邊我也不太確定為甚麼會這樣，推測有可能是data太少造成的誤差

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | **λ** | RMSE (training data) | RMSE (validation) |
| 全部feature | 0.1 | 5.710510 | 5.040508 |
| 0.01 | 5.710504 | 5.040527 |
| 0.001 | 5.710503 | 5.040528 |
| 0.0001 | 5.710503 | 5.040529 |
| Only PM2.5 | 0.1 | 6.135208 | 5.537837 |
| 0.01 | 6.135208 | 5.537822 |
| 0.001 | 6.135208 | 5.537820 |
| 0.0001 | 6.135208 | 5.537820 |

**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)為一存量 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 ？請寫下算式並選出正確答案。(其中XTX為invertible)**

1. **(XTX)XTy**
2. **(XTX)-0XTy**
3. **(XTX)-1XTy**
4. **(XTX)-2XTy**

ANS : **(XTX)-1XTy**

Proof :

**最小化loss function 🡪**

**🡪 🡪**

**故得證!!**