說明：請各位使用此template進行Report撰寫，如果想要用其他排版模式也請註明題號以及題目內容（請勿擅自更改題號），最後上傳至github前，請務必轉成PDF檔，並且命名為report.pdf，否則將不予計分。

----------------------------------------閱讀完以上文字請刪除------------------------------------

學號：B02123456 系級： 電機四 姓名：法拉利

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)

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

自己先算，如果有一天還有剩下的quota，再丟 Kaggle

1. learning rate 10-5，太大的話無法 converge，太小的話train很慢（我沒用optimizer）。epoch 數越大，validation error、training error 皆降低(尚未converge)，我先做到 300，希望換 model 降error的效果會比硬train 還好。Batch size 似乎在 20 的時候最好。離最佳解（closed form）還有一段距離，hyperparameter 很難tune到最好（batch, epoch, lr…）
2. learning rate 10-4，epoch=50。Batch size 似乎在 10 的時候最好。Closed form 會發現 error 提升，但是實際 train 出來的會比較(1)還好，而且更接近 closed form solution，可能因為 feature 比較少？

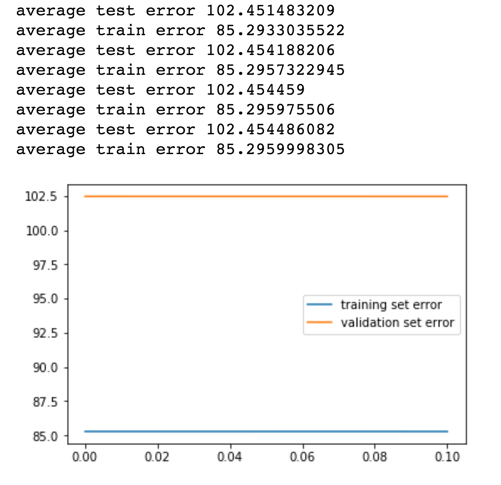
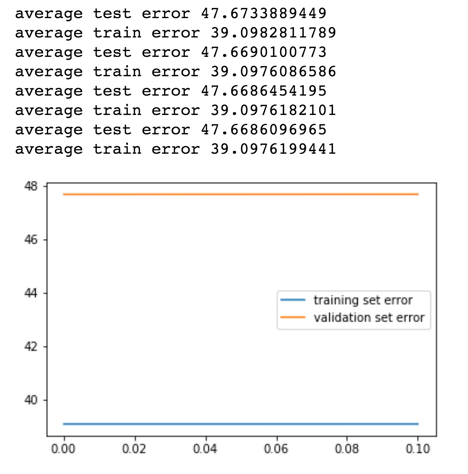
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

(1) error會下降，train速度提升，爽。但是hyperparam改變，batch現在改成30較佳。但還是沒有單獨9小時PM2.5當feature好。

(2) error一樣會下降，train速度提升，爽。Hyperparam沒改變 。

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

全部 feature 都是一次方，有意義嗎…



All 9 features, PM2.5 only

Both consider all 9 hours

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] 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。

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