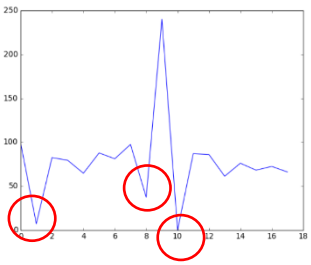
學號：R05921037 系級： 電機碩一 姓名：陳冠廷

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

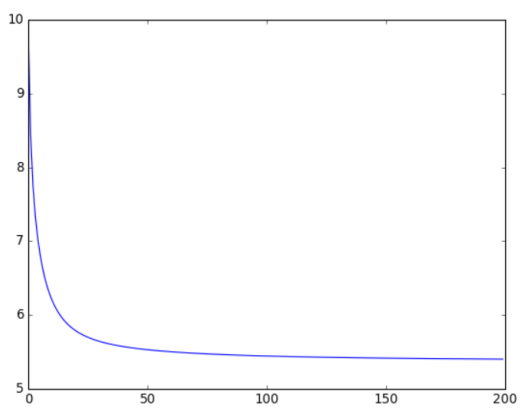
答：

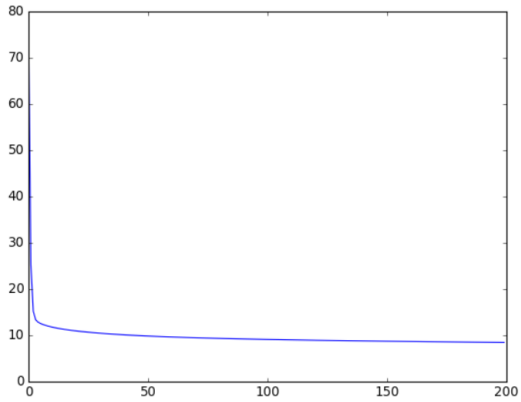
用 open 的方式先把每一行 row 抓下來，再把多餘的 column、row 濾掉(例如 train.csv 的第一行)，然後先用 split(,240)先切成 240 個 24*18 的矩陣，再用 for 迴圈把 numpy 矩陣用[i:i+9]決定 162 個 feature，最後先用 coorelation 做出右圖，把最小的三個 feature 濾掉，形成 135 個 feature 的一個 input data。



2.請作圖比較不同訓練資料量對於 PM2.5 預測準確率的影響

答：

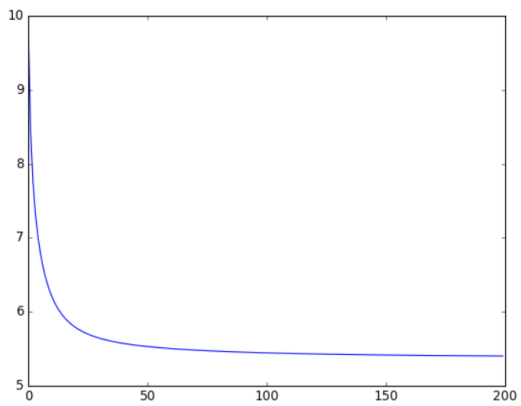


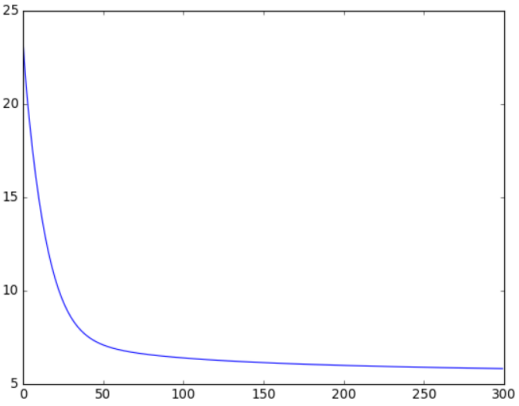


左圖為用 Stochastic Gradient Descent，右圖為一次丟進所有 data，可以看到兩個極端例子，batch size=1 時下降的幅度可以到 5.5(RMSE)才趨於平穩，而 batch size 是所有 data 時降到 9.8 多就不會往下掉，我想這是由於當一筆一筆資料去辨認時，路徑會比較明確，但若是整個資料丟進去，每個 weight 的變量會是所有資料所算出 gradient descent 的合，這樣會導致一些資訊因為加加減減彼此抵消而消失。

3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

答：





左圖為二次式(mx^2+nx+b)又圖為一次式(nx+b)，可以看出二次式明顯的比較快到最低點，且二次式收斂的位置較低，二次式收斂到 5.513 而一次式收斂到 5.920，而為何兩個其實分數都不錯的原因我想是我已經把三個不相干的資訊去除的因素。

4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響

答：

以這次實驗的感覺，我認為是在降低 gradient descent 的下降速度並且不讓學習 overfitting，但其參數加上原本的 learning rate 有兩個參數變量，所以後來就用 Adagrad 的方式讓學習不會 overfitting。

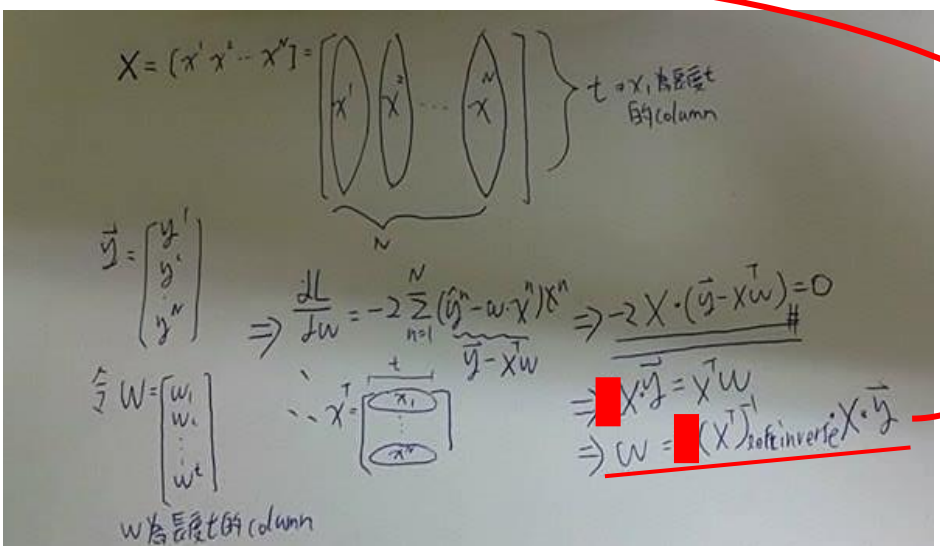
左圖是用 adagrad 右圖沒用，可以看到不只 loss 下降慢而且 validation error 也很大，我覺得可能是因為每次 iteration 的速率沒下降導致震盪讓學習曲線無法很直接地往正確的方向走。

```
50  L: 5.63601727574 delta(loss): 124.358541048
VALIDATION: 5.29993195795
100  L: 5.52529364693 delta(loss): 34.6609888958
VALIDATION: 5.35127207619
150  L: 5.4842166007 delta(loss): 16.2006602455
VALIDATION: 5.36176961291
200  L: 5.4627106665 delta(loss): 9.36311119935
VALIDATION: 5.36382824481
250  L: 5.44941960024 delta(loss): 6.13902934517
VALIDATION: 5.3644269942
```

```
50  L: 9.12092398237 delta(loss): 1126.69826472
VALIDATION: 21.797816276
100  L: 8.30189703092 delta(loss): 514.299522922
VALIDATION: 19.9341389703
150  L: 7.81795453945 delta(loss): 328.52313674
VALIDATION: 18.3859598593
200  L: 7.46602972884 delta(loss): 241.296000651
VALIDATION: 17.0203000919
250  L: 7.1868327414 delta(loss): 188.908012829
VALIDATION: 15.7938859521
```

5. 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$\sum_{n=1}^{N} (y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \ldots x^N]$ 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，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：　　W= ● [(Xtranspose)left inverse]● X● $\vec{y}$



上面不小心加上了 2 所以用紅色蓋掉