

Homework 1 Report - PM2.5 Prediction

學號：r06921077 系級：電機碩一 姓名：黃詩凱

1. (1%) 請分別使用每筆data9小時內所有feature的一次項 (含bias項) 以及每筆data9小時內PM2.5的一次項 (含bias項) 進行training，比較並討論這兩種模型的root mean-square error (根據kaggle上的public/private score)。

	Private score	Public score	本機RME
只用PM2.5一次項	9.69186	9.55812	23.340141
全部feature	8.77186	9.13797	22.724894

只使用PM2.5的解集合會被包含在使用全部feature的解集合中
所以理論上用Gradient Descent，只要不要走到Local minimum的話，使用全部feature的結果不一定會比只使用PM2.5的一次項差
當然前提是他們都有收斂的狀況下，像我這個測試，跑40000次，只有用PM2.5的收斂完成，使用全部feature的可能因為資料量大，40000次還能繼續往下走
所以以實驗結果來看，算是蠻合理的

2. (2%) 請分別使用至少四種不同數值的learning rate進行training (其他參數需一致)，作圖並且討論其收斂過程。

初始參數 and 條件：

只採用PM2.5過去的值來Train

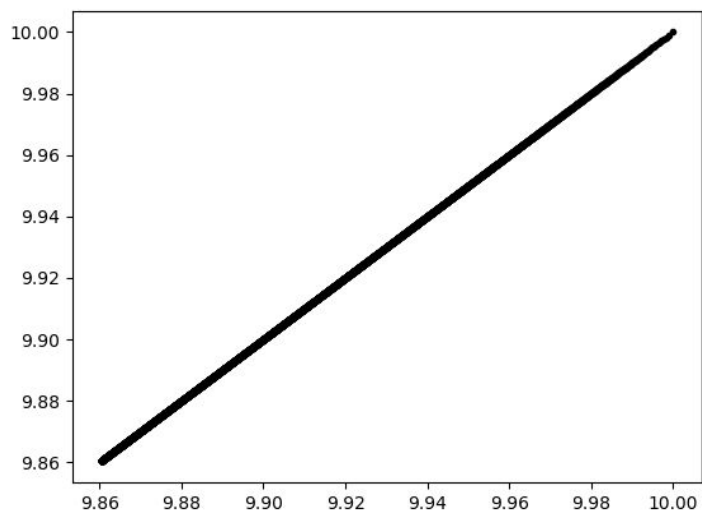
有用到Adagrad

iteration: 5000

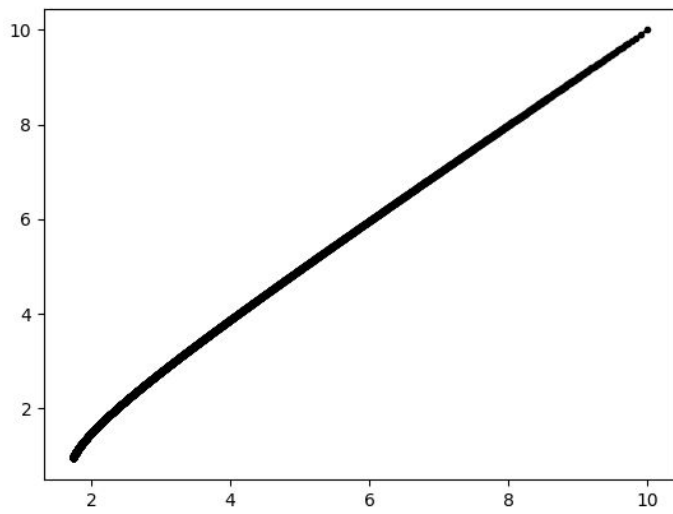
init w = 10

init b = 10

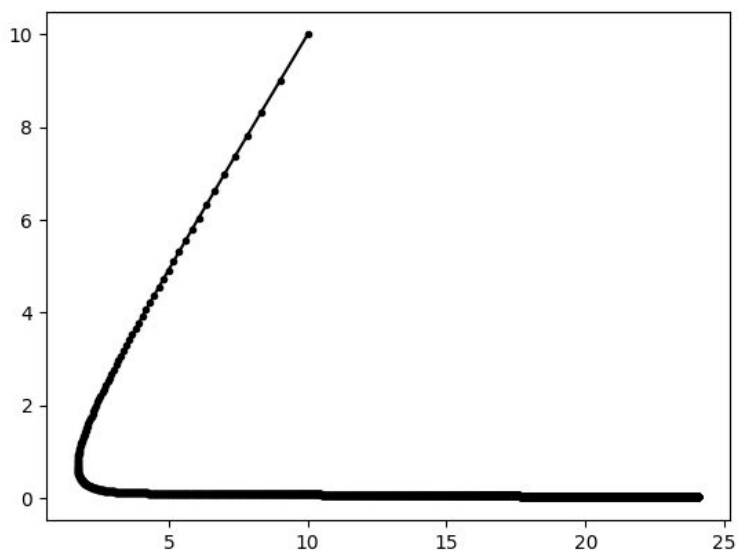
learning rate: 0.001



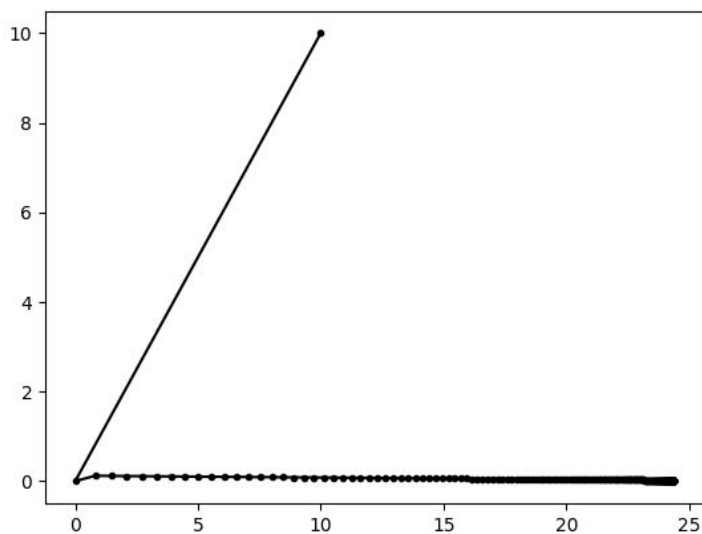
learning rate: 0.1



learning rate: 1



learning rate: 10



由以上可以觀察到learning rate = 0.001 和 0.1 時，步伐太小，iteration也很少

所以沒有像後面兩個一樣往右邊走，誤差就很大

以上面四種情況來說，比較好的狀況應該是learning rate=1時，步伐剛剛好，比較容易快速走到低點

3. (1%) 請分別使用至少四種不同數值的regularization parameter λ 進行training (其他參數需一至)，討論其root mean-square error (根據kaggle上的public/private score)。

learning rate: 10

iteration: 40000

λ	private score	public score	本機RME
1	8.45120	8.49661	22.835019
5	8.45393	8.50287	22.835572
10	8.45731	8.51029	22.836267
100	8.50074	8.58898	22.846076

可以觀察到 λ 愈大，error也愈大

就我推測可能是因為Regularization是用來解決overfitting的問題

它能盡量壓低參數的大小，讓函數更平滑

但是我的方法只使用了一次項，overfitting的問題不太會碰到

再加上其他參數不變，然後loss加上一項之後會比較慢收斂，所以error才愈來愈大

4. (1%) 請這次作業你的best_hw1.sh是如何實作的？(e.g. 有無對Data做任何 Preprocessing？Features的選用有無任何考量？訓練相關參數的選用有無任何依據？)

hw1_best主要是由Gradient Descent + Adagrad實作

主要是依據助教給的Sample code參考修改的

Preprocessing部分只有把NR改成0

Data Feature只挑了CO, NO, NO2, PM10, PM2.5, RAINFALL

這部分我是自己手動試，發現這幾個feature結果都還不錯，如果加其他feature或是刪掉其中幾個feature表現則不一定好。

自己一開始有嘗試只使用PM2.5下去Train，但結果沒有很好，所以就找了一些理論上較相關的feature拿來嘗試、比較。

也有嘗試Cross validation，可以比較容易判斷哪個Model比較好，減少Overfitting的發生