

**(a) Motivation and approach**

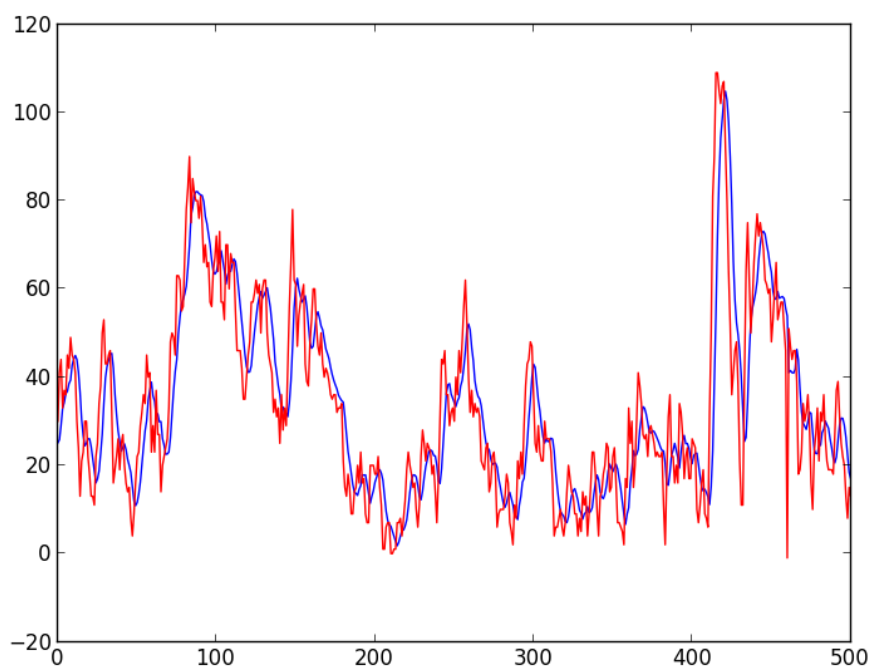
本次作業中要求我們以 linear regression by gradient descent 的方法 train 2014 年豐原站每個月前 20 天 PM2.5 的數值，並 test 後 10 天某連續 9 小時後第 10 小時的結果，所以我把這  $12(\text{月}) * 20(\text{天}) * 24(\text{小時})$  筆資料，每 9 小時一取，形成約 5000 組 dataset。而我的模組是以  $y = w_1x_1 + w_2x_2 + \dots + w_9x_9 + b$  的方程式 training，所以我要 train 的參數數目一共有 9 個。在一開始的時候，我的 loss function 並沒有隨著時間下降，反而是以非常快的速度上升，於是我調整我的 learning rate 從 0.01 到 0.0001，果然 loss function 就開始下降了，一開始因為一直收斂不到我要求的條件，所以我開始設定 training 次數，收斂的速度大約從 200 次之後開始下降，後來嘗試了 Adaptive learning rate 的方式，但因為原本的收斂速度就是從後面才開始下降，而這個方法的 learning rate 也是會隨著次數變小，所以最後效果並不顯著。

在最後要加入 regularization term，我的  $\lambda$  從 1, 10, 100, 1000 試了四組數字，loss function 也隨著數字的提升變得越來越大，所以判斷我的曲線應該是有變的比較平滑。

但由於加入 regularization term 會使得上傳到 Kaggle 上的分數

無法到達 Baseline 所以在程式中最後還是把這個項目給拿掉了。

### (b) Output



這是我取前 500 筆 dataset，train 出來結果，紅色為實際值，藍色為估計值。

### (c) Discussion

在這次的 Homework 中，實作了 gradient descent 的方法，知道了 learning rate 對整個 training 的重要性，尤其是起始值特別重要，會影響整個結果是否收斂，我的模擬設定在 10000 次，但其實在約 1000 次的時候，收斂幅度已經很小了。而在 kaggle 上的分數也只在 5.8 左右，在這部分應該是可以再加強，例如取更高次方，或是加入更多 feature 等等。