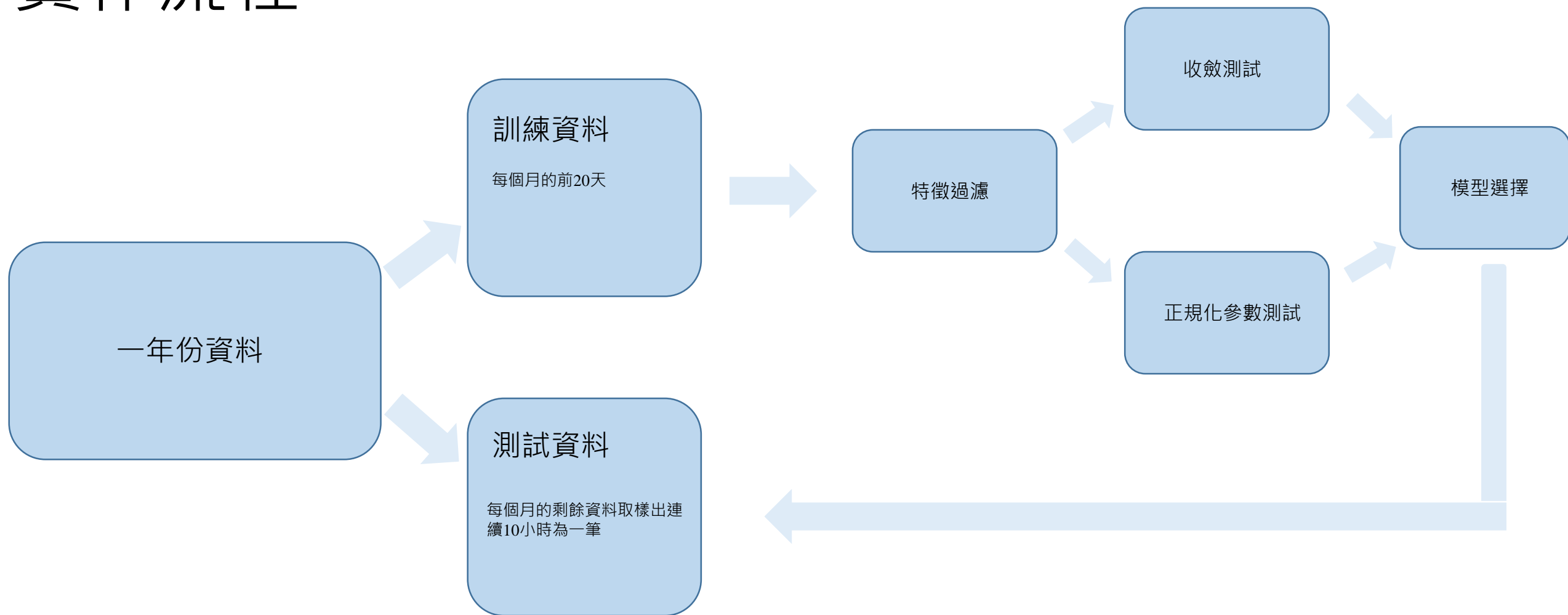


空氣品質預測

- 利用中央氣象局豐原站所提供空氣品質資料抓取1整年份資料
預測建模來預測空氣中的pm2.5數值
- 模型透過輸入前9個小時pm2.5的數值來預測當下pm2.5的數值
我們的模型表現誤差範圍RMSE = 7.29

實作流程



技術細節

- 訓練資料在本專案中將後面20%資料切分為驗證資料。
- 由於我們特徵選擇皆為pm2.5僅為不同小時的區別，此時每個特徵的scale是差不多的，因此這裡我們沒有做feature scaling
- 由於我們採用Linear model，因此objective function為convex function，不必擔心saddle points及local min的問題。
- 優化算法採用adaptive gradient，可自我調整的梯度算法可以節省模型測試時間，如Figure 1，模型於550左右個迭代穩定收斂

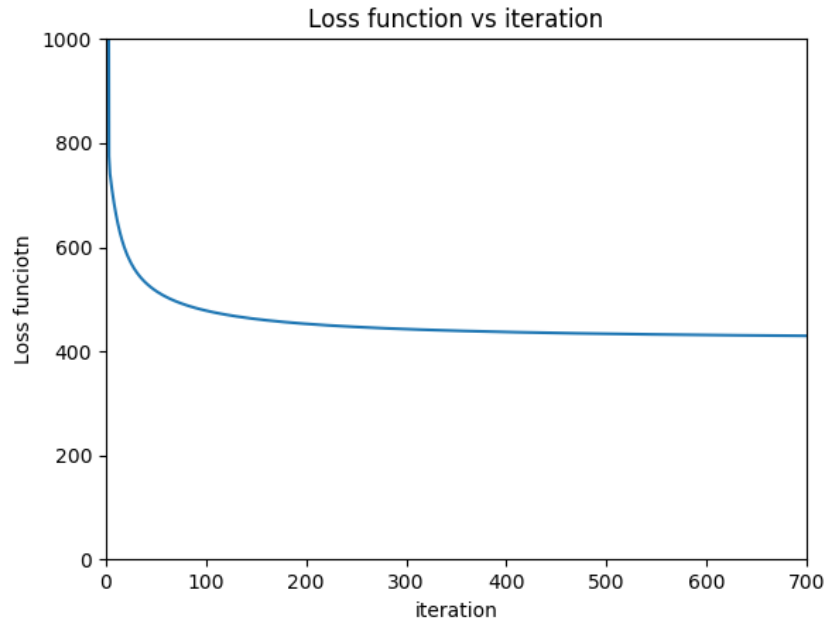


Figure 1

Iteration表示算法迭代的次數，Loss function為SE，初始學習率 $\eta = 1$

技術細節

- 經過反覆測試，使用L1 Regularizer 或是 L2 Regularizer所得到的 E_{val} 數值差僅在小數點後4位，因此我們認定效果是差不多的，其中 E_{val} 為驗證資料集誤差，誤差衡量為MAE。
- 模型選擇如下Figure 2 所示， E_{val} 的最小值落在 $lambda = 0.0001$ 其值為 4.636。
- 模型選擇完成後，我們使用未切分過的訓練資料重新訓練一個模型，如此可以使誤差在小一些。

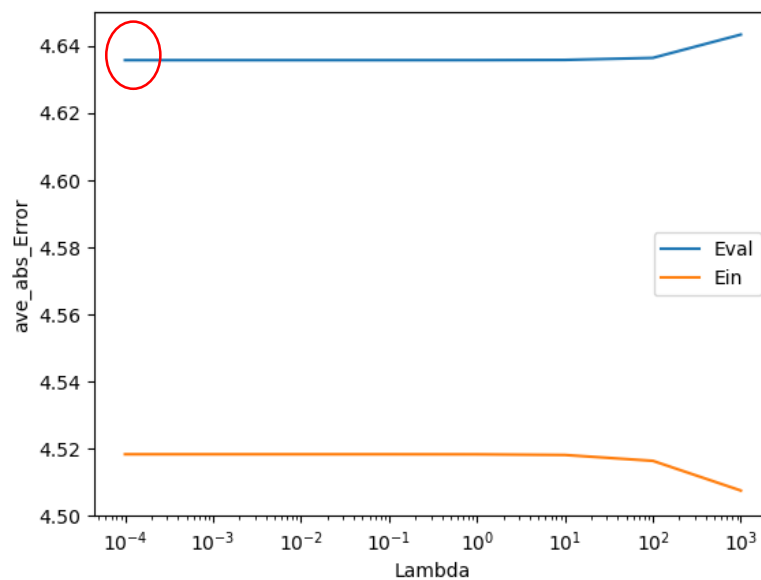


Figure 2

x軸為正規化參數lambda，y軸為誤差衡量，採用MAE。