A063529 吳燕萍

1.只有PM2.5的時間序列迴歸模型

使用前六期的PM2.5資料預測下一期的PM2.5

迴歸方程式如下:

|  |  |
| --- | --- |
| 變數名稱 | 係數 |
| 截距項 | 10.629672551778064 |
| LAG1 | 0.14592256 |
| LAG2 | 0.03935985 |
| LAG3 | 0.04966078 |
| LAG4 | -0.04638318 |
| LAG5 | 0.0303036 |
| LAG6 | 0.09881222 |

2.只有PM2.5的隨機森林

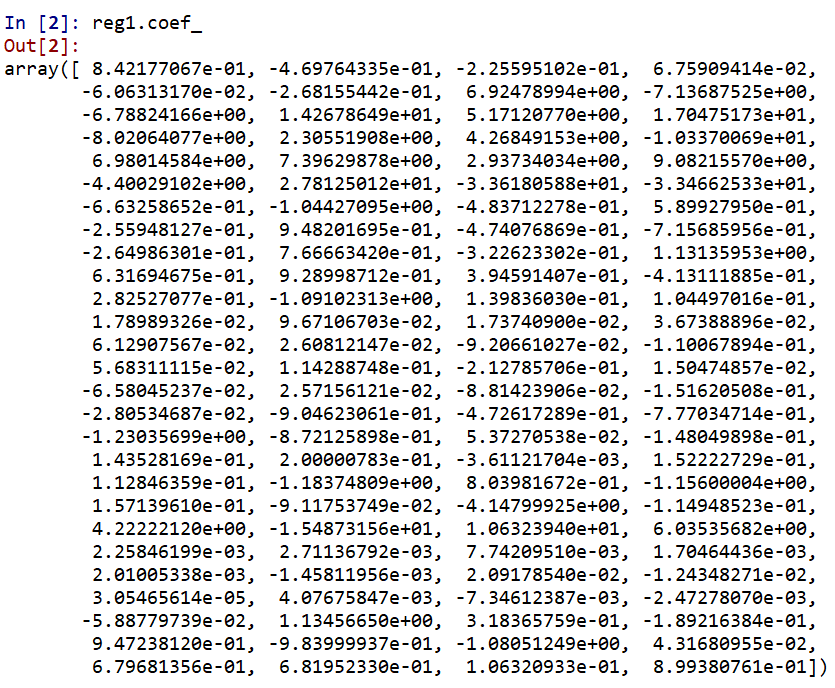
參數設定:500棵樹、leaf node至少包含五筆資料、每次隨機選取2個特徵值做best split(因為一共有6個落後期的資料，因為是迴歸問題，因此最好每次隨機篩選p/3=6/3=2(round down)個特徵值)，並且設定隨機種子為0

3.所有18種屬性的時間序列迴歸模型

使用前六期的資料去預測下一期的PM2.5，其中前六期的資料包含所有的觀測項目，因此一共有18\*6個特徵。

以下截圖是它的迴歸係數(從第一個屬性的LAG1開始，接續是第一個屬性的LAG2…)

截距項: -88.7364257352892



4.所有18種屬性的隨機森林

參數設定:500棵樹、leaf node至少包含五筆資料、每次隨機選取2個特徵值做best split(因為一共有6個落後期的資料，因為是迴歸問題，因此最好每次隨機篩選p/3=6/3=2(round down)個特徵值)，並且設定隨機種子為0

* **模型績效(MAE)** 用測試資料與四種模型預測的結果作比對

|  |  |
| --- | --- |
|  | MAE |
| 只有PM2.5的時間序列迴歸模型 | 9.936985320586759\* |
| 只有PM2.5的隨機森林 | 9.975633760850334 |
| 所有18種屬性的時間序列迴歸模型 | 11.624869392277475 |
| 所有18種屬性的隨機森林 | 9.777989700481417 |

由結果我們可以看出，只有利用PM2.5的屬性的前六期資料去預測當期PM2.5的時間序列模型，若以MAE來衡量模型的表現，它是表現得最好的。