

使用不同資料處理方式及學習模型預測早上 6 點鐘的 PM2.5

Github: <https://github.com/HUAN178/PM2.5-analysis>

垂直型：

```
In [22]: PM25_inside = PM25[['0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '21', '22', '23']] #取出PM2.5資料 共2208筆
PM25
```

	id	日期	測站	測項	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	10	2018-10-01	新竹	PM2.5	9.0	17.0	15.0	17.0	18.0	13.0	14.0	19.0	20.0	19.0	21.0	22.0	18.0	19.0	24.0	31.0	33.0	34.0	38.0	42.0	40.0	38.0	36.0	29.0
1	28	2018-10-02	新竹	PM2.5	28.0	31.0	27.0	21.0	16.0	19.0	12.0	11.0	16.0	13.0	14.0	455.0	18.0	17.0	19.0	20.0	16.0	17.0	17.0	15.0	14.0	13.0	15.0	13.0
2	46	2018-10-03	新竹	PM2.5	9.0	8.0	10.0	11.0	11.0	13.0	17.0	13.0	10.0	10.0	10.0	13.0	16.0	13.0	11.0	11.0	14.0	14.0	11.0	15.0	21.0	17.0	13.0	17.0
3	64	2018-10-04	新竹	PM2.5	9.0	17.0	19.0	17.0	14.0	17.0	18.0	17.0	16.0	14.0	13.0	14.0	14.0	14.0	17.0	19.0	18.0	17.0	18.0	17.0	15.0	16.0	13.0	12.0
4	82	2018-10-05	新竹	PM2.5	13.0	15.0	16.0	17.0	19.0	19.0	19.0	19.0	21.0	22.0	23.0	23.0	21.0	24.0	24.0	22.0	19.0	21.0	24.0	23.0	28.0	31.0	27.0	27.0
5	100	2018-10-06	新竹	PM2.5	27.0	27.0	29.0	29.0	24.0	20.0	23.0	26.0	24.0	19.0	19.0	22.0	21.0	19.0	20.0	22.0	22.0	26.0	28.0	26.0	25.0	22.0	21.0	23.0

資料處理方式：訓練機器學習時利用垂直資料(過去每天固定時段的 PM2.5 值)

作為特徵值預測未來每天 6 點鐘的 PM2.5 值。

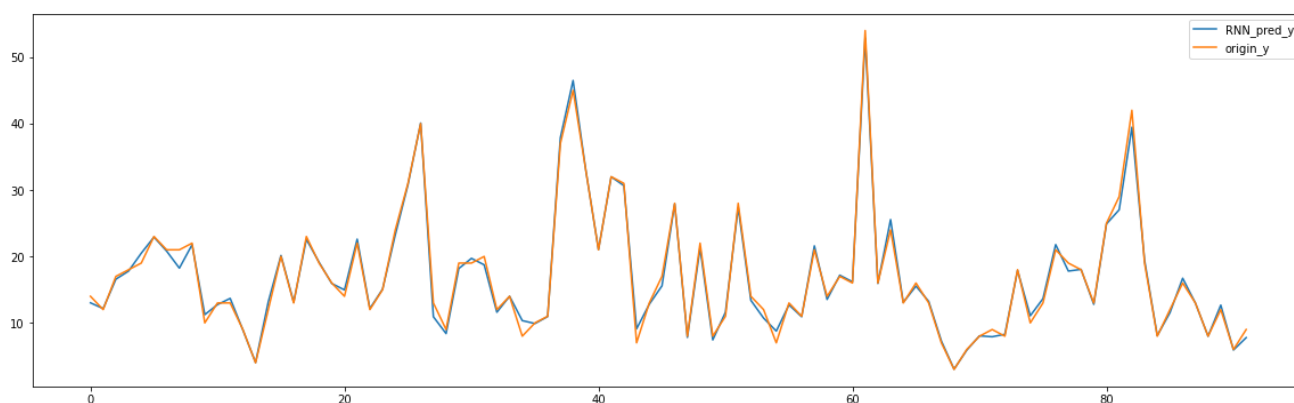
特色：樣本稀缺時使用，且訓練機器學習時只考量幾天前特定時段的特徵值，未將全部連續時段的值納入考量。

結果：

1.LinearRegression：Test R Square: 0.94

Train R Square:0.91

2.RNN：預測未來每天 6 點鐘 PM2.5 值的 MSE 值為 25.5



連續水平型：

```
In [22]: PM25_inside = PM25[['0','1','2','3','4','5','6','7','8','9','10','11','12','13',  
                             '14','15','16','17','18','19','20','21','22','23']] #取出PM2.5資料 共2208筆  
PM25
```

	id	日期	測站	測項	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	10	2018-10-01	新竹	PM2.5	19.0	17.0	15.0	17.0	18.0	13.0	14.0	19.0	20.0	19.0	21.0	22.0	18.0	19.0	24.0	31.0	33.0	34.0	38.0	42.0	40.0	38.0	36.0	29.0
1	28	2018-10-02	新竹	PM2.5	28.0	31.0	27.0	21.0	16.0	19.0	12.0	11.0	16.0	13.0	14.0	455.0	18.0	17.0	19.0	20.0	16.0	17.0	17.0	15.0	14.0	13.0	15.0	13.0
2	46	2018-10-03	新竹	PM2.5	9.0	8.0	10.0	11.0	11.0	13.0	17.0	13.0	10.0	10.0	10.0	13.0	16.0	13.0	11.0	11.0	14.0	14.0	11.0	15.0	21.0	17.0	13.0	17.0
3	64	2018-10-04	新竹	PM2.5	19.0	17.0	19.0	17.0	14.0	17.0	18.0	17.0	16.0	14.0	13.0	14.0	14.0	14.0	17.0	19.0	18.0	17.0	18.0	17.0	15.0	16.0	13.0	12.0
4	82	2018-10-05	新竹	PM2.5	13.0	15.0	16.0	17.0	19.0	19.0	19.0	19.0	21.0	22.0	23.0	23.0	21.0	24.0	24.0	22.0	19.0	21.0	24.0	23.0	28.0	31.0	27.0	27.0
5	100	2018-10-06	新竹	PM2.5	27.0	27.0	29.0	29.0	24.0	20.0	23.0	26.0	24.0	19.0	19.0	22.0	21.0	19.0	20.0	22.0	22.0	26.0	28.0	26.0	25.0	22.0	21.0	23.0

資料處理方式：去除極端值後訓練機器連續利用每 6 小時當特徵值預測下 1 小時的 PM2.5 數值，以利預測未來每天 6 點鐘的 PM2.5 值。

特色：訓練資料數量眾多且具連續性，能以大量的連續特徵值訓練機器學習。又因樣本數眾多，可設計較多模型種類分析預測。

結果：

1. LinearRegression：整體 Test R Square: 0.9131

整體 Train R Square: 0.0443

2. RNN：預測未來每天 6 點鐘 PM2.5 值的 MSE 值為 29.44，整體測試的 MSE 值為 14.63。

