

Assignment 1: PM 2.5 Prediction

Due Mar 20, 2018, 11:59 pm

發布時間：2017/02/25 23:00
即時細懸浮微粒指標

Task - Predict PM2.5

本次作業的資料是從中央氣象局網站下載的真實觀測資料，希望大家利用linear regression或其他方法預測PM2.5的數值。



請點擊左方測站位置或

所屬單位：環保署 ▾

地區：中部 ▾ > 忠明 ▾ 查詢

發布時間：2017-02-25 23:00:00

忠明 (一般站)

AQI 1

細懸浮微粒指標 低

PM_{2.5}($\mu\text{g}/\text{m}^3$) 移動平均 10

細懸浮微粒 小時濃度 18

單位： $\mu\text{g}/\text{m}^3$ ，微克/立方公尺

ND：未檢出(表示數據低於偵測極限2微克/立方公尺)

PM_{2.5}移動平均值計算方式： $0.5 \times$ 前12小時平均 + $0.5 \times$ 前4小時平均(前4小時3筆有效，前12小時8筆有效)

低	低	低	中	中	中	高	高	高	非常高
1	2	3	4	5	6	7	8	9	10

● 監測車

⊗ 設備維護(測站例行維護、儀器異常維修、監測數據不足)

Dataset

- ❖ 本次作業使用豐原站的觀測記錄，分成train set跟test set, train set是豐原站每個月的前20天所有資料。test set則是從豐原站剩下的資料中取樣出來。
- ❖ **train.csv**: 每個月前20天的完整資料。
- ❖ **test.csv**: 從剩下的資料當中取樣出連續的10小時為一筆，前九小時的所有觀測數據當作feature，第十小時的PM2.5當作answer。一共取出240筆不重複的test data，請根據feature預測這240筆的PM2.5。

Training Data ([link](#))

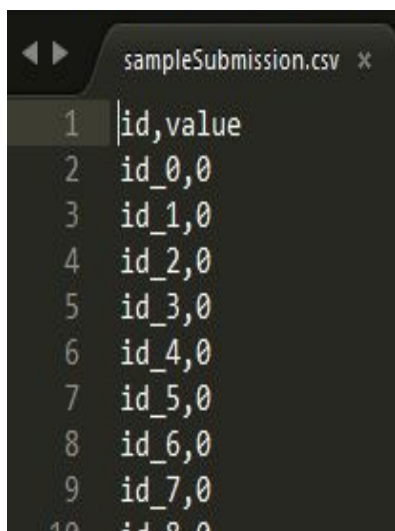
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	日期	測站	測項	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
2	2014/1/1	豐原	AMB_TEM	14	14	14	13	12	12	12	12	15	17	20	22	22	22	22	
3	2014/1/1	豐原	CH4	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	
4	2014/1/1	豐原	CO	0.51	0.41	0.39	0.37	0.35	0.3	0.37	0.47	0.78	0.74	0.59	0.52	0.41	0.4	0.37	
5	2014/1/1	豐原	NMHC	0.2	0.15	0.13	0.12	0.11	0.06	0.1	0.13	0.26	0.23	0.2	0.18	0.12	0.11	0.1	
6	2014/1/1	豐原	NO	0.9	0.6	0.5	1.7	1.8	1.5	1.9	2.2	6.6	7.9	4.2	2.9	3.4	3	2.5	
7	2014/1/1	豐原	NO2	16	9.2	8.2	6.9	6.8	3.8	6.9	7.8	15	21	14	11	14	12	11	
8	2014/1/1	豐原	NOx	17	9.8	8.7	8.6	8.5	5.3	8.8	9.9	22	29	18	14	17	15	14	
9	2014/1/1	豐原	O3	16	30	27	23	24	28	24	22	21	29	44	58	50	57	65	
10	2014/1/1	豐原	PM10	56	50	48	35	25	12	4	2	11	38	56	64	56	57	52	
11	2014/1/1	豐原	PM2.5	26	39	36	35	31	28	25	20	19	30	41	44	33	37	36	
12	2014/1/1	豐原	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	N
13	2014/1/1	豐原	RH	77	68	67	74	72	73	74	73	66	56	45	37	40	42	47	
14	2014/1/1	豐原	SO2	1.8	2	1.7	1.6	1.9	1.4	1.5	1.6	5.1	15	4.5	2.7	3.5	3.6	3.9	
15	2014/1/1	豐原	THC	2	2	2	1.9	1.9	1.8	1.9	1.9	2.1	2	2	2	1.9	1.9	1.9	
16	2014/1/1	豐原	WD_HR	37	80	57	76	110	106	101	104	124	46	241	280	297	305	307	
17	2014/1/1	豐原	WIND_DIR	35	79	2.4	55	94	116	106	94	232	153	283	269	290	316	313	
18	2014/1/1	豐原	WIND_SPEED	1.4	1.8	1	0.6	1.7	2.5	2.5	2	0.6	0.8	1.6	1.9	2.1	3.3	2.5	
19	2014/1/1	豐原	WS_HR	0.5	0.9	0.6	0.3	0.6	1.9	2	2	0.5	0.3	0.8	1.2	2	2.6	2.1	
20	2014/1/2	豐原	AMB_TEM	16	15	15	14	14	15	16	16	17	20	22	23	24	24	24	

Testing Data ([sample test](#), [sample outputs](#))

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id_0	AMB_TEM	15	14	14	13	13	13	13	13	12		
2	id_0	CH4	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8		
3	id_0	CO	0.36	0.35	0.34	0.33	0.33	0.34	0.34	0.37	0.42		
4	id_0	NMHC	0.11	0.09	0.09	0.1	0.1	0.1	0.1	0.11	0.12		
5	id_0	NO	0.6	0.4	0.3	0.3	0.3	0.7	0.8	0.8	0.9		
6	id_0	NO2	9.3	7.1	6.1	5.7	5.5	5.3	5.5	7.1	7.5		
7	id_0	NOx	9.9	7.5	6.4	5.9	5.8	6	6.2	7.8	8.4		
8	id_0	O3	36	44	45	44	44	44	43	40	38		
9	id_0	PM10	51	51	31	40	34	51	42	36	30		
10	id_0	PM2.5	27	13	24	29	41	30	29	27	28		
11	id_0	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR		
12	id_0	RH	75	71	71	73	74	74	74	74	74		
13	id_0	SO2	1.2	1.2	1.2	1.6	1.5	1.5	1.5	1.6	1.6		
14	id_0	THC	1.9	1.8	1.8	1.9	1.9	1.9	1.9	1.9	1.9		
15	id_0	WD_HR	116	114	112	109	111	104	107	108	104		
16	id_0	WIND_DIR	115	113	105	102	106	106	112	113	106		
17	id_0	WIND_SPEED	2.6	2.2	2	1.9	2.4	2.4	2.5	2.8	2		
18	id_0	WS_HR	2.1	2.4	2.2	1.9	2.3	2.3	2.5	2.5	2.3		
19	id_1	AMB_TEM	12	12	12	13	14	15	14	14	13		
20	id_1	CH4	1.8	1.8	1.9	1.9	1.8	1.8	1.8	1.8	1.8		

Prediction Results

- ❖ 預測test set中的240筆PM2.5, 上傳至e-learning。
 - 格式為csv (filename: **predictions.csv**)
 - 第一行必須是 id,value
 - 第二行開始, 每行分別為id及預測數值, 以逗點分開
- ❖ 範例格式:



```
sampleSubmission.csv x
1 id,value
2 id_0,0
3 id_1,0
4 id_2,0
5 id_3,0
6 id_4,0
7 id_5,0
8 id_6,0
9 id_7,0
10 id_8,0
```

Requirements

- ❖ Only **Python 3.5+** available
- ❖ Implement linear regression with **Gradient Descent**.
- ❖ Packages for gradient descent are not allowed. Only numpy, scipy and pandas are allowed. Standard libraries are okay. But `numpy.linalg.lstsq` is forbidden
- ❖ **Versions:**
Python 3.5+
numpy 1.13

Submission

1. hw1.py, other files required to run your code, and the trained model.
Zip all of your files into a single filename.zip, where filename is your student ID.
2. You must train your model yourself. We will load your trained model and test it on the test data.
3. To train your model: `hw1.py --mode train --file train.csv`
4. To test your model: `hw1.py --mode test --file test.csv`

Your Grades

1. Your grades are based on the performance on the test data
2. Your model will be compared with others and will receive a rank. The rank and the test error will determine your grades.
3. More on the grades later.