

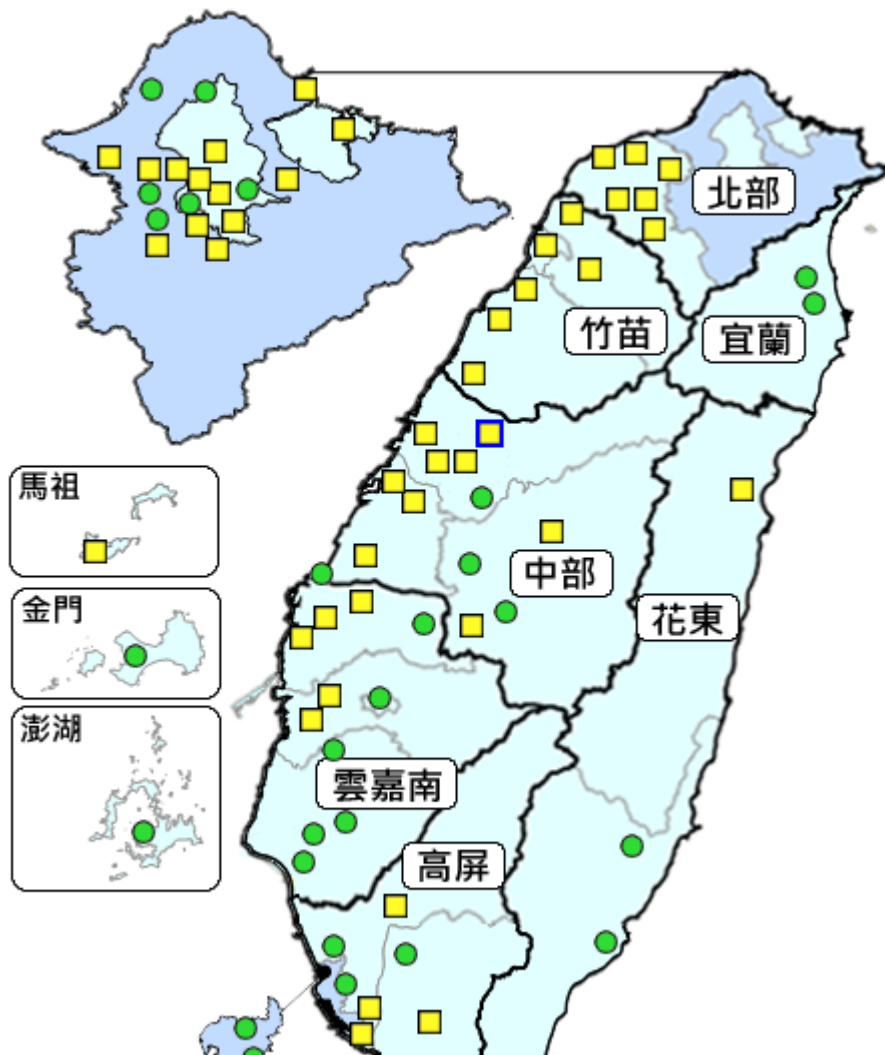
HW1 – PM2.5 Prediction

Task Description

預測豐原站在下一個小時
會觀測到的 PM2.5

地區： >

發布時間： 2017-09-29 08:00:00




豐原 (一般站)

AQI **64**
空氣品質指標 **普通**

O ₃ (ppb) 臭氧	8小時 移動平均	17
	小時 濃度	14
PM _{2.5} (µg/m ³) ◎ 細懸浮微粒	移動 平均	21
	小時 濃度	27
PM ₁₀ (µg/m ³) 懸浮微粒	移動 平均	35
	小時 濃度	44

Task Description

- 預測 A 年 B 月 C 日 N 時的 PM2.5
 - 每個時間點以一個 ID 表示，共 240 個時間點
 - 評比標準：預測值和實際值的平方誤差平均值
- 預測根據：前九小時的所有觀測數據
 - A 年 B 月 C 日 N - 1 時的 PM2.5, CH4, NO, NO2, O3 ...
 - A 年 B 月 C 日 N - 2 時的 PM2.5, CH4, NO, NO2, O3 ...
 -
 - A 年 B 月 C 日 N - 9 時的 PM2.5, CH4, NO, NO2, O3 ...


$$f(\text{前九小時的所有觀測數據}) = \text{A 年 B 月 C 日 N 時的 PM2.5}$$

Testing Data

	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		

Three Steps for Machine Learning

- Step 1. Define you function set (Model)

y : A 年 B 月 C 日 N 時的 **PM2.5**

$x_{K,M}$: A 年 B 月 C 日 N - K 時的 M 觀測值

M = **PM2.5**, **NO**, **O3**, CH4, NO2 ... (總共 18 種)

How to deal with RAINFALL = NR?

$$y = b + w_{1,\text{PM}} \cdot x_{1,\text{PM}} + w_{2,\text{PM}} \cdot x_{2,\text{PM}} + \dots + w_{9,\text{PM}} \cdot x_{9,\text{PM}}$$

$$\begin{aligned} y = & b + w_{1,\text{PM}} \cdot x_{1,\text{PM}} + w_{2,\text{PM}} \cdot x_{2,\text{PM}} + \dots + w_{5,\text{PM}} \cdot x_{5,\text{PM}} \\ & + w_{1,\text{NO}} \cdot x_{1,\text{NO}} + w_{2,\text{NO}} \cdot x_{2,\text{NO}} + \dots + w_{5,\text{NO}} \cdot x_{5,\text{NO}} \\ & + w_{1,\text{O3}} \cdot x_{1,\text{O3}} + w_{2,\text{O3}} \cdot x_{2,\text{O3}} + \dots + w_{5,\text{O3}} \cdot x_{5,\text{O3}} \end{aligned}$$

$$\dots + \lambda \text{ 所有 } w \text{ 的平方和相加} \quad \lambda = ?$$

Three Steps for Machine Learning

- Step 2. Define your loss function based on training data

	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													3	2.5
7	2014/1/1	豐原	NO2	16													12	11
8	2014/1/1	豐原	NOx	17													15	14
9	2014/1/1	豐原	O3	16													57	65
10	2014/1/1	豐原	PM10	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
13	2014/1/1	豐原	RH	77	68	67	74	72	73								42	47
14	2014/1/1	豐原	SO2	1.8	2	1.7	1.6	1.9	1.4								3.6	3.9
15	2014/1/1	豐原	THC	2	2	2	1.9	1.9	1.8								1.9	1.9
16	2014/1/1	豐原	WD_HR	37	80	57	76	110	106								105	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

Training Data 來自每個月的前 20 天
Testing Data 來自每個月的後 10 天

請自己從 Training Data
切 Validation Data

Three Steps for Machine Learning

- Step 3. Find the best function
 - Please use Gradient Descent (of course, there are other approaches).

Tips

- The error surface of linear regression is convex.
 - No matter the feature you use.
- You can **CHECK** your results by looking for the closed-form solution.
 - This is why we choose it as HW1.

