

Homework 1 Report - PM2.5 Prediction

學號：b04501095 系級：土木三 姓名：黃平瑋

1. (1%) 請分別使用每筆data9小時內所有feature的一次項 (含bias項) 以及每筆data9小時內PM2.5的一次項 (含bias項) 進行training，比較並討論這兩種模型的root mean-square error (根據kaggle上的public/private score)。

optimizer = adam

regularization = 0.1

iteration = 10000

	參數數量	training error	testing error
全部數據都取	163	6.844860	7.041545
只取PM2.5	10	11.4608	7.73521

* testing error = (public score + private score)/2

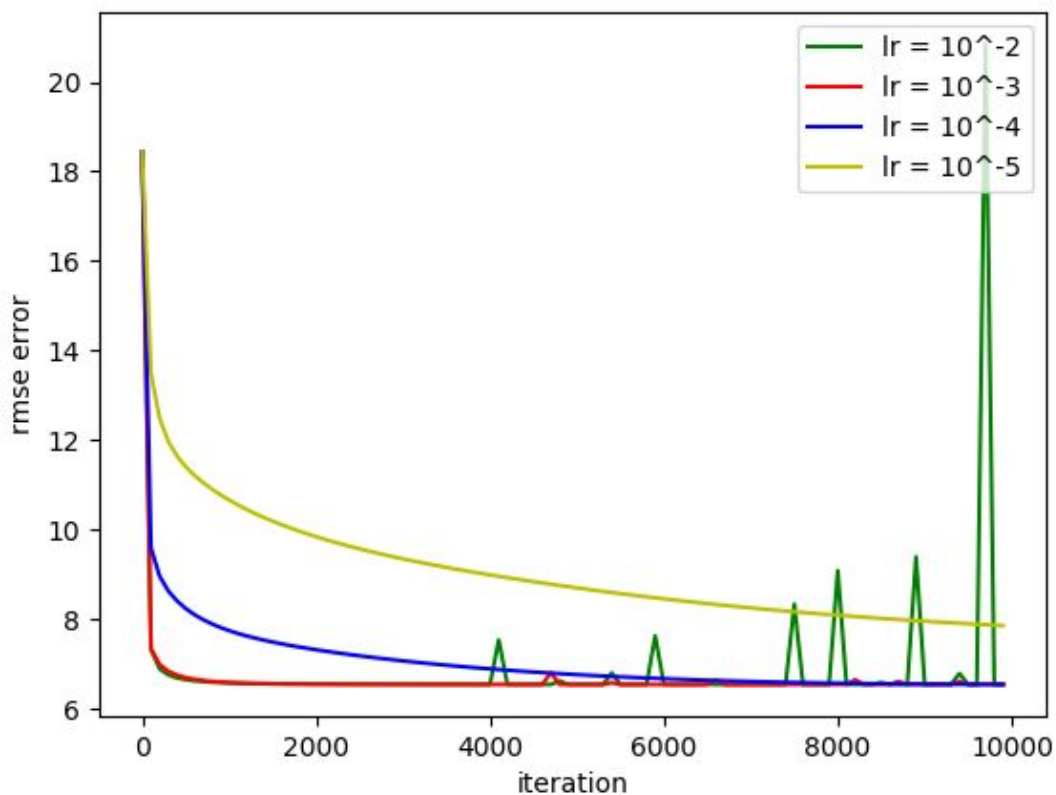
由上圖的表格可以清楚知道，取所有feature的結果，不論是在training set 或是 testing set 表現都比只取PM2.5的結果還要好，雖然PM2.5的主要的決定因素，但是其他的因素，如風速 雨量 污染物 可能都會造成PM2.5數值的浮動，因此將更多feature考慮進來，可以有效的降低error。

2. (2%) 請分別使用至少四種不同數值的learning rate進行training (其他參數需一致)，作圖並且討論其收斂過程。

optimizer = adam

regularization = 0.1

iteration = 10000



因為所有的feature我都有先做了normalization，所以learning rate 不能調太大，不然會看到error出現明顯的震盪，如learning rate = 0.01的情況。但如果learning rate調太小，如learning rate = 10^{-5} 的情況，就可能10000次iteration結束後，卻還是沒有抵達minimum

3. (1%) 請分別使用至少四種不同數值的regularization parameter λ 進行training (其他參數需一致，討論其root mean-square error (根據kaggle上的public/private score))。

regularization parameter λ	training error	public error	private error	testing error
1000	7.105681	7.72462	7.42655	7.57559
100	6.594201	6.89762	7.74589	7.32176
10	6.539953	6.73026	7.68595	7.24763
1	6.539745	6.73222	7.69079	7.211505
0.1	6.539761	6.80862	7.68664	7.208105

在我的實做中，當regularization parameter λ 很小時似乎沒有造成太大的影響，只有在 λ 值漸漸變大時，會使testing set error變大，可能是較大的 λ ，模型會比較不重視預測落差反而是極力地想要使曲線變得更平滑

4. (1%) 請這次作業你的best_hw1.sh是如何實作的？(e.g. 有無對Data做任何Preprocessing？Features的選用有無任何考量？訓練相關參數的選用有無任何依據？

我在觀察training data時發現資料似乎有點錯誤，有些時間的所有污染物都是0，如圖一而有些時間的數值會比前後幾小時的值高出很多，如圖二。此外，也有一些時間點，污染物的數值為負的。

會有以上幾種情況出現，我認為是在觀測時機器發生問題，或是數據有遺漏或紀錄錯誤的情況，但如果就直接把這些資料刪掉，可能會使我的training set變太小，增加overfitting的機會，為了補足這些有瑕疵的數據，我會將它們以各污染物的平均值替代。

此外所有的資料都有經過標準化，讓不同feature間的比較能較為客觀

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
2014/4/11	日	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
2014/4/11	日	RH	75	77	77	77	81	79	79	78	73	66	61	54	52	56	54	50	52	53	68	75	77	77	76	73
2014/4/11	日	SO2	5	7.6	5.4	2.8	1.9	2.2	3.8	6.8	7	5.5	4.6	4.3	4.4	4.2	5.5	8.9	14	13	15	8.3	6.1	8	8.6	5.6
2014/4/11	日	THC	2.5	2.3	2.6	2.5	2.3	2.1	2.3	2.5	2.3	2.2	2	1.9	1.8	1.8	1.8	1.9	1.9	2	2.5	2.2	2.2	2.2	2.2	2.3
2014/4/11	日	WD_HR	254	225	180	200	265	239	231	243	232	257	259	236	251	259	264	267	265	262	242	212	215	209	149	154
2014/4/11	日	WIND_DIREC	230	166	171	264	280	237	259	242	234	262	264	266	269	258	261	272	258	260	229	197	253	174	115	183
2014/4/11	日	WIND_SPEED	0.7	0.8	0.6	1.8	0.8	0.7	1.2	1.3	1.3	1.9	1.6	1.7	3.3	4.2	3.7	3.7	2.8	2	1.7	1.6	1.2	0.8	1.2	0.8
2014/4/11	日	WS_HR	1.1	0.6	0.6	0.3	0.9	0.5	0.8	1.1	1.1	1.2	1.4	1	2.1	3.3	3.5	3.6	3.3	2.6	1.7	1.2	1	0.7	0.6	0.6
2014/4/12	日	AMB_TEMP	22	22	21	21	22	22	22	24	27	30	32	33	34	34	0	0	0	0	0	0	0	26	26	26
2014/4/12	日	CH4	1.9	2	2.1	2.2	2	2.1	1.9	1.9	1.8	1.7	1.7	1.7	1.7	1.6	0	0	0	0	0	0	0	-0.2	-0.2	-0.2
2014/4/12	日	CO	0.8	0.7	0.68	0.66	0.48	0.54	0.63	0.68	0.54	0.42	0.36	0.42	0.34	0.27	0	0	0	0	0	0	0	0.45	0.34	0.36
2014/4/12	日	NMHC	0.35	0.31	0.59	0.6	0.4	0.37	0.36	0.28	0.2	0.2	0.14	0.15	0.08	0.06	0	0	0	0	0	0	0	0.02	0.02	0.02
2014/4/12	日	NO	1.7	1.6	2.1	2.6	1	1.4	5	7.6	6.4	2.8	1.4	1.2	1.1	1.1	0	0	0	0	0	0	0	0.6	0.3	0.5
2014/4/12	日	NO2	28	26	27	26	18	21	22	22	20	14	9.8	12	9.7	8.1	0	0	0	0	0	0	0	-0.6	11	12
2014/4/12	日	NOx	30	28	30	29	19	22	27	30	26	17	11	13	11	9.2	0	0	0	0	0	0	0	0	12	13
2014/4/12	日	O3	12	9	4.9	4.9	16	10	13	19	31	61	71	91	78	69	0	0	0	0	0	0	0	1	1	27
2014/4/12	日	PM10	100	92	88	91	88	89	83	88	84	87	79	77	70	66	0	0	0	0	0	0	0	66	85	71
2014/4/12	日	PM2.5	56	53	46	47	49	47	42	37	41	37	45	42	48	36	0	0	0	0	0	0	0	-3	11	12
2014/4/12	日	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	0	0	0	0	0	0	0	0	NR	NR
2014/4/12	日	RH	74	76	79	79	73	75	73	68	62	52	46	45	46	47	0	0	0	0	0	0	0	69	70	69
2014/4/12	日	SO2	3.8	2.8	3.5	4	3	2.1	3.9	3	2.5	3.5	3.1	4.5	3.8	2.8	0	0	0	0	0	0	0	-0.3	6.3	5.4
2014/4/12	日	THC	2.3	2.4	2.7	2.8	2.4	2.4	2.3	2.2	2	1.9	1.8	1.9	1.8	1.7	0	0	0	0	0	0	0	-0.2	-0.2	-0.2
2014/4/12	日	WD_HR	213	237	235	206	192	175	161	242	224	210	248	268	267	269	0	0	0	0	0	0	0	214	192	182
2014/4/12	日	WIND_DIREC	193	229	254	143	251	162	192	219	230	303	262	263	270	270	0	0	0	0	0	0	0	213	187	171
2014/4/12	日	WIND_SPEED	0.6	0.5	1.1	0.7	1.1	0.8	0.6	1.3	1.7	1.3	2.4	2.2	3.5	2.8	0	0	0	0	0	0	0	1.2	1.2	1.2
2014/4/12	日	WS_HR	0.5	0.6	0.7	0.4	0.4	0.8	0.7	1	1.1	0.9	1.3	2.4	3.2	2.6	0	0	0	0	0	0	0	1.2	1	0.9
2014/4/13	日	AMB_TEMP	26	26	25	25	25	25	25	27	29	31	33	34	34	34	34	33	33	32	30	28	27	26	25	25
2014/4/13	日	CH4	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.3	-0.3	-0.3	-0.2	1.7	1.6	1.6	1.6	1.6	1.6	1.6	1.7	1.7	1.7	1.7
2014/4/13	日	CO	0.33	0.33	0.33	0.34	0.33	0.34	0.46	0.49	0.4	0.36	0.36	0.28	0.25	0.22	0.22	0.23	0.27	0.36	0.45	0.3	0.36	0.31	0.33	0.33

圖一

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
2014/4/1	日	CH4	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.8	1.8	1.8	1.8	1.8	1.7	1.7	1.7	1.8	1.7	1.7
2014/4/1	日	CO	0.41	0.39	0.4	0.38	0.39	0.41	0.43	0.87	0.92	0.78	0.67	0.69	0.68	0.6	0.52	0.69	0.55	0.7
2014/4/1	日	NMHC	0.09	0.09	0.05	0.04	0.05	0.07	0.07	0.22	0.26	0.28	0.3	0.33	0.31	0.22	0.16	0.32	0.23	0.29
2014/4/1	日	NO	0.2	0.5	1.3	0.9	1.2	0.9	1.3	4.6	4.8	4.8	4.1	5.3	4.3	3.7	2.9	2.6	2.4	2.2
2014/4/1	日	NO2	8.6	7.9	6.6	5.9	7.2	9.5	13	25	28	26	25	26	23	18	15	26	21	28
2014/4/1	日	NOx	8.8	8.4	7.8	6.8	8.4	10	14	29	32	31	29	31	27	22	18	29	23	30
2014/4/1	日	O3	43	43	46	48	45	38	34	24	24	27	28	29	39	47	53	39	34	23
2014/4/1	日	PM10	82	80	72	56	30	22	32	38	37	34	38	42	47	49	47	48	49	48
2014/4/1	日	PM2.5	42	40	44	36	18	14	14	19	18	16	17	17	15	23	21	24	20	24
2014/4/1	日	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
2014/4/1	日	RH	81	81	81	80	78	78	78	76	73	73	72	70	66	62	62	64	69	72
2014/4/1	日	SO2	0.7	2	1.8	2	2.2	2.5	3.5	4	4.3	3.7	3.3	3.6	3.9	3.3	3	3.1	2.7	3.2
2014/4/1	日	THC	1.8	1.8	1.8	1.8	1.8	1.8	1.8	2	2	2	2.1	2.1	2	1.9	1.9	2.1	2	2
2014/4/1	日	WD_HR	290	18	323	29	328	5.7	355	298	312	306	336	300	312	318	317	290	304	7.5
2014/4/1	日	WIND_DIRECT	272	309	2.8	45	299	328	358	300	301	324	11	280	341	294	8.9	309	311	323
2014/4/1	日	WIND_SPEED	1.2	2.1	2.3	1.5	1	1.4	1.7	1.4	1.7	1.7	2.1	1.4	1.9	2.1	1.3	2.5	2.4	2.6
2014/4/1	日	WS_HR	0.7	0.7	1	1	0.7	0.5	0.7	0.9	1	1.1	0.7	0.8	0.9	0.9	0.7	1.2	0.9	0.7
2014/4/2	日	AMB_TEMP	19	19	19	19	19	18	18	19	21	23	25	23	19	19	20	23	23	21
2014/4/2	日	CH4	1.8	1.9	1.8	1.8	2	2.1	2	2	2	1.8	1.7	1.8	1.8	1.7	1.7	1.7	1.8	1.7
2014/4/2	日	CO	0.53	0.51	0.42	0.35	0.38	0.4	0.49	0.84	0.93	0.8	0.54	0.57	0.65	0.5	0.52	0.68	0.82	0.67
2014/4/2	日	NMHC	0.2	0.22	0.14	0.1	0.19	0.2	0.26	0.36	0.38	0.3	0.22	0.37	0.41	0.72	0.85	0.68	0.43	0.27
2014/4/2	日	NO	1.2	1.6	1.2	0.7	0.6	1	3.7	14	17	8.1	4	2.5	5.1	-0.3	3.5	4.3	3.7	1.8
2014/4/2	日	NO2	20	21	17	13	16	18	25	28	33	27	18	20	35	0	23	27	30	23
2014/4/2	日	NOx	21	23	18	14	16	19	29	42	50	35	22	23	40	-1.3	26	31	34	25
2014/4/2	日	O3	17	16	21	26	18	16	12	9.8	14	25	40	29	19	0	0	28	35	39
2014/4/2	日	PM10	34	31	34	22	19	27	30	34	29	37	0	36	42	0	39	37	39	47
2014/4/2	日	PM2.5	21	22	26	15	12	3	10	13	19	16	0	631	5	0	12	15	7	10
2014/4/2	日	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	1	1	0.4	0.6	NR	NR	NR

圖二