

Generative Classification Models

The Homework 3 report from NTU102-1 [DMIR](#) course

by NTU [Michael Hsu](#)

Data Pre-process

1. 新增欄位 `index` : 原本資料的排序 (= sort by `week_index` and `group`) 。
2. 排序與篩選 `week_index` + `week_return1` :

week_index	group	alpha	beta_mkt	beta_hml	beta_smb	sigma	week_retu	week_retu	week_return13
1	1	0.116891	2.051794	1.593424	3.217965	0.271982	0.892801	0.900921	0.895643
2	1	0.033864	3.903705	3.1536	5.176311	0.099684	0.886809	0.875238	0.886578
3	1	0.006197	1.664317	1.515463	-1.03395	0.03316			
4	1	0.010429	2.791942	0.975734	3.03268	0.06887			
5	1	0.005181	2.382342	3.754779	1.944673	0.03496			
6	1	0.015829	0.125074	1.126783	0.802143	0.10894			
7	1	0.005273	0.725278	1.851199	0.988507	0.04601			
8	1	0.00652	2.724019	2.074522	2.951118	0.0583			
9	1	0.003305	2.718963	2.832244	2.839584	0.0531			
10	1	0.004364	-0.39638	-0.34048	-0.87334	0.08243			
11	1	0.002783	-1.41999	-1.57716	-0.85298	0.05523			
12	1	0.000811	0.923352	1.454901	0.613547	0.02675			
13	1	0.000959	1.833449	-0.98287	0.300169	0.03246			
14	1	0.00182	-0.02956	-0.35929	0.337405	0.07101			
15	1	0.000308	-0.58653	-0.66232	-0.32575	0.03662			
16	1	0.000151	0.997433	0.676193	-1.03792	0.020894	1.031823	1.043708	1.043407
17	1	-0.00076	0.311303	0.281951	0.201965	0.030383	1.026217	1.011423	1.006739
18	1	-0.00077	0.141854	-0.07859	-0.05794	0.029051	0.982732	1.00561	0.998664
19	1	-0.00123	-0.69946	-1.10135	-0.3587	0.031414	0.963968	0.983456	0.983645
20	1	-0.00091	0.122074	0.079034	0.029426	0.009608	1.00565	1.00773	1.004218
21	1	0.00177	0.21365	0.35276	0.13106	0.018603	1.014684	1.0008	0.999721

3. 定義分類標籤 :
 - 1. 新增欄位 `index_sort` : 根據上一個步驟後的排序。
 - 1. 新增欄位 `index_sort % 30` : `mod(左邊, 30)`
 - 1. 給予分類標籤
`class` : `=IF((左邊>0)*(左邊<=6),"1","0")` 前六個為 1, 剩下二十四個為 0。
4. 新增欄位 `random_sort` : 最後依據這個欄位 `=RAND()` 來做 10-fold classification。
5. 最後整理資料為 `data/ldpa30_train use.csv`
 - 1. 剩下 feature

- `alpha` 、 `beta_mkt` 、 `beta_hml` 、 `beta_smb` 、 `sigma`
- 分類的標籤 `class`
- 以及目前的隨機排序依據，作為切割十份用，產生新的 `new_index` 。

new_index	alpha	beta_mkt	beta_hml	beta_smb	sigma	class
1	-0.0065066	3.048392	4.07517039	0.9617558	0.01808332	0
2	0.00057031	0.32789225	0.4956953	0.59191614	0.06036168	0
3	0.00236977	1.44659396	-1.835827	2.10289862	0.01848847	0
4	0.00019669	1.51353161	1.08480384	1.94350839	0.01727014	0
5	-0.0042819	1.45120471	1.86734677	0.59260946	0.02066149	0

如何執行

R cmd:

```
> source("/path_to/generative_classification_model.r")
```

example: (可用拖曳方式取得路徑)

```
> source("/Users/michaelhsu/Dropbox/15.\ 碩一上課業/02.\ DMIR\
```

結果 (10-fold-validation)

```

accuracy precision    recall  F-measure
bin1  0.6927928 0.3098160 0.4654378 0.3720074
bin2  0.6891892 0.3108974 0.4273128 0.3599258
bin3  0.6720721 0.2916667 0.4375000 0.3500000
bin4  0.6711712 0.2586751 0.3867925 0.3100189
bin5  0.6837838 0.3123028 0.4267241 0.3606557
bin6  0.6738739 0.2296820 0.3110048 0.2642276
bin7  0.6954955 0.3030303 0.4072398 0.3474903
bin8  0.6711712 0.2743902 0.4147465 0.3302752
bin9  0.6774775 0.3021148 0.4405286 0.3584229
bin10 0.6900901 0.3302469 0.4572650 0.3835125
      accuracy precision    recall  F-measure
mean 0.681711712 0.29228222 0.41745519 0.34365365
sd    0.009647542 0.02995744 0.04395451 0.03470257
```

Source code

https://github.com/evenchange4/102-1_DMIR_Hw3_Generative-Classification-Models