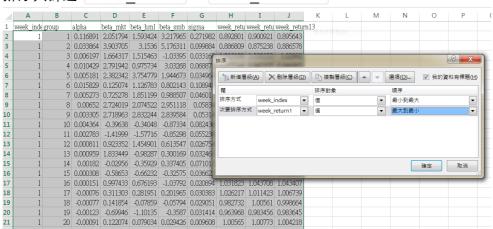
# **Generative Classification Models**

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 :



- 3. 定義分類標簽:
  - 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
  - 。 剩下 feature

```
alpha \ beta_mkt \ beta_hml \ beta_smb \ sigma
```

- 。 分類的標簽 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
II C W	THUEN

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
    0.009647542 0.02995744 0.04395451 0.03470257
sd
```