AIC による変数選択
竹内・鈴木関数のビジネス利用へ

M.Shimura
JCD02773@nifty.ne.jp

2005 年 12 月 22 日

1　AIC で変数を選ぶ

2005 年 8 月の蓼科でのサマーセミナーで竹内が変数を選択する場合に、多くのクライテリアで判別する方法を提示し、翌月に、鈴木が AIC のみを用いる簡単な 2 の関数を示した。

現実の問題として ESRI(内閣府経済社会総合研究所) が毎月発表する景気動向指数 (Leading Coinsist Lagged) の月次指標の項目が 32 ある。これから重相関に用いる指標を 10 項目程度選びだそうとすると、

${}_{10}C_{32}$

10!32 =64512240 の組み合わせが出来る。竹内鈴木関数で選択しなければ収拾がつかないし、計算時間から AIC のみとする。

1.1　竹内・鈴木関数のテスト

サンプルデータ (R) から 9 変数を選択する場合の 9 個の組み合わせと AIC 値 (ソート後・最初の 10 個)

竹内・鈴木関数はデータ X は横長の形で受け付けるようになっている。縦長のデータを用いるときは (|:) で rotate しておく。

find_best0 は AIC の小さい順に並べる。

```
10{. find_best0 Y1 compare 9;R
```

```
+------------------------+-------+
|1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+------------------------+-------+
|1 1 0 1 0 1 1 1 1 0 1 0 1|209.542|
+------------------------+-------+
|1 1 0 1 1 1 1 1 0 0 1 0 1|209.602|
+------------------------+-------+
|1 1 0 1 0 1 1 1 0 1 1 0 1|209.603|
+------------------------+-------+
|1 1 1 1 0 1 1 1 0 0 1 0 1|209.614|
+------------------------+-------+
|1 1 0 1 0 1 0 1 1 0 1 1 1|209.667|
+------------------------+-------+
|1 1 1 1 0 1 0 1 1 0 1 0 1|209.691|
+------------------------+-------+
|1 1 0 1 1 1 0 1 1 0 1 0 1|209.705|
+------------------------+-------+
|1 1 0 1 0 1 0 1 0 1 1 1 1|209.714|
+------------------------+-------+
|1 1 0 1 0 1 0 1 1 1 1 0 1|209.718|
+------------------------+-------+
```

find_best は組み合わせの範囲を指定して、各組毎の、AIC の最小値を表示する。find_best0 を見ても、AIC の値は多少うつろうが、並べてみて、0 の谷の深いところを外していけば、大綱では概ね妥当な選択が出来ている。

組み合わせは、${}_{m}C_{n}$ の n が大きいと out_of_memory を頻発する。予選リーグを行い、L C LG 毎に相関の少ないものを先に外していくと計算量がずっと少なくなって、妥当な結果が得られる。

7 変数から 12 変数の場合の AIC 最小の組み合わせを選ぶ。

```
  (7;12)  find_best  Y1; R
+--+------------------------+-------+
|7 |1 1 0 1 0 1 0 1 0 0 1 0 1|209.881|
```

```
+--+------------------------+-------+
|8 |1 1 0 1 0 1 1 1 0 0 1 0 1|209.615|
+--+------------------------+-------+
|9 |1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+--+------------------------+-------+
|10|1 1 0 1 0 1 1 1 0 1 1 1 1|209.395|
+--+------------------------+-------+
|11|1 1 0 1 1 1 1 1 0 1 1 1 1|209.36 |
+--+------------------------+-------+
|12|1 1 0 1 1 1 1 1 1 1 1 1 1|209.357|
+--+------------------------+-------+
```

2 Refelence

Script

```
NB. J.Takeuchi discuss best fit using various criteria
NB. At Tateshina seminor Aug 2005,
NB. G.Suzuki simplified using AIC
NB. next 2 script was written by G.Suzuki Sept/2005
   stat_reg=:3 :0
regb=:[%.1:,.] NB.regression coefficient
regp=:(1:,.])+/ .*regb NB.predicted value
regq=:[:+/[:*:[-regp NB.sum of residuale
regcd=:100"_*1:-regq%[:+/[:*:(]-+/%#)@[
mat=:[:%.(|:+/ .*])@(1:,.]) NB.inverse of data matrix
resvar=:regq%[:-/[:$1:,.] NB.residual varience
regt=:regb%[:%:resvar*[:(<1 0)&|:mat@]
mll=:>:@^.@((o.2)"_*regq%#@[)*#@[%_2: NB.Mll
regaic=:+:@(1:+#@(1:,:]))-2:*mll
'program set of regression model'
)
```

```
NB.    stat_reg ''
NB. program set of regression model
NB.    $ R=:A,B,C,D,E,F,G,H,I,J,K,L,:M
NB. 13 87

compare=:4 :0
s=.((>{.y.)=+/"1 t)#t=.#:i.2^#y=.>{:y.
r=.,:(k{s);x.regaic|:(k=.0){u=.s#y
while. k<<:#s
  do. r=.r,(k{s);x.regaic|:(k=.k+1){u
end.
)

NB. hereinafter 2 script is written by M.Shimura
NB. combinient for using in actual

find_best0=:3 : 0
NB. Usage u. Y1 compare 9;R
TMP0=: y.
(/:  {:"1 TMP0){ TMP0
)

find_best =: 4 : 0
NB. Usage: (7;12)  find_best  Y1; R
Y=: ;{. y.
ZONE=:  ({. ; x.) + i. -/ |. ; x.
ANS=: 0;0
COUNTER=: 0
while. COUNTER < # ZONE do.
X=: (COUNTER { ZONE);;{: y.
TMP1=: {. find_best0 Y compare X
ANS=: ANS,TMP1
COUNTER=. >: COUNTER
```

```
end.
({@> ZONE),.((,(# ZONE),2)$  2 }. ANS)
)
```