

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

ПРАКТИЧНЕ МАШИННЕ НАВЧАННЯ.

BOOTSTRAP

Звіт

до лабораторної роботи №6

з курсу «Методи і засоби комп'ютерного навчання»

Виконав:

студент гр. СПКм-12

Сергієнко В.Р.

Прийняв

Пукач А.І.

Львів 2014

Мета: оволодіти методами статистичного аналізу, використовуючи бібліотеку *bootstrap*, в системі R.

Короткі теоретичні відомості

bootstrap (x, nboot, theta ,..., func = NULL)

Аргументи:

x – вектор, що містить дані.

nboot – число необхідних *bootstrap* наборів даних.

theta – функція, яка потрібна *bootstrap*. Приймає *x* в якості аргумента, також може мати додаткові аргументи.

інші – додаткові аргументи, що передаються функції *theta*.

func – додатковий аргумент, що визначає бажану функцію розподілу. Якщо аргумент *func* визначений, повертається оцінка стандартної помилки для даної величини.

Повертаються значення:

thetastar – *nboot bootstrap* значень *theta*.

func.thetastar – функціонал *func bootstrap* розподілу *thetastar*, якщо *func* визначений.

jack.boot.val – jackknife після завантаження значення для *func*, якщо *func* визначений.

jack.boot.se – jackknife після завантаження оцінка стандартної помилки *func*, якщо *func* визначений.

call – рядок, що представляє собою “розібраний” виклик функції

Індивідуальне завдання

3. Завантажити набір даних *law* і обчислити кореляцію між середнім балом, набраним учнями однієї школи на тесті з правознавства (*LSAT*), і їхнім середнім балом з усіх предметів (*GPA*). Оцінити стандартну помилку коефіцієнта кореляції, використовуючи *bootstrap*. Обчислити коефіцієнт кореляції на наборі даних *law82* і порівняти отримані результати.

Виконання завдання

```
library(bootstrap)
xdata=law
perc05 <- function(x)
{
  quantile(x, .05)
}
n=15
law_corr=cor(xdata)
law_corr
LSAT      GPA
LSAT 1.0000000 0.7763745
GPA 0.7763745 1.0000000
theta=function(x,xdata)
{
```

```

        cor(xdata[x,1],xdata[x,2])
    }
law_cor_boot=bootstrap(1:n,20,theta,xdata)
law_cor_boot$thetastar
[1] 0.6322209 0.3635928 0.7416313 0.8555713 0.8394284 0.7882003 0.8869730
[8] 0.5782711 0.8451423 0.8629898 0.9300517 0.8653002 0.9330735 0.6529697
[15] 0.8836331 0.7925031 0.7888651 0.8112843 0.3621745 0.2585020
ydata=law82
y=subset(t,select=-School)
theta=function(y,ydata)
{
    cor(ydata[y,1],ydata[y,2])
}
law82_corr=cor(ydata)
law82_corr
      School      LSAT      GPA
School  1.00000000 -0.05527412 -0.0967626
LSAT    -0.05527412  1.00000000  0.7599979
GPA     -0.09676260  0.75999786  1.0000000
82_cor_boot=bootstrap(1:n,20,theta,ydata)
law82_cor_boot=bootstrap(1:n,20,theta,ydata)
law82_cor_boot$thetastar
[1] 0.03487786 0.13810613 0.10838315 0.43535477 -0.17120061 0.46370518
[7] 0.05901148 0.38489922 0.32565884 -0.55934572 0.60597357 -0.06939080
[13] 0.28859084 0.39311760 0.82222905 0.25985906 -0.01309579 0.33332296
[19] -0.55754796 0.18924712

```

Висновок:

Виконуючи лабораторну роботу, навчився проводити статистичний аналіз використовуючи бібліотеку *bootstrap*, в системі R.