Regresión múltiple: estudio empírico del orden de la multiplicación de matrices.

Taller MAT III GIUN. 1920

Contents

Enunciado

Cargamos los datos, los transformamos en un `data frame` de dos variables `n` y `seconds`

```
#cargamos el objeto del working directory actual
load("data1000.Robj")
class(data1000)
```

```
## [1] "matrix"
```

```
#no muestro la extructura pues es muy larga
#str(data1000)
#head(data1000)
data=as.data.frame(matrix(unlist(data1000),ncol=2,byrow=TRUE))
head(data)
```

```
##    V1    V2
## 1  10 0.000
## 2 110 0.001
## 3 210 0.003
## 4 310 0.008
## 5 410 0.013
## 6 510 0.026
```

```
#ponemos nombres
names(data)=c("n","seconds")
#eliminamos las filas con seconds=0
data=data[data$seconds!=0,]
head(data)
```

```
##     n seconds
## 2 110   0.001
## 3 210   0.003
## 4 310   0.008
## 5 410   0.013
```

```
## 6 510   0.026
## 7 610   0.045
```

Estudio del orden por regresión múltiple

En el taller anterior estudiamos si los modelos de orden del algoritmo de multiplicación de matrices se ajustan a un modelo lineal a uno exponencial o a uno potencial y recordamos que sabemos que la multiplicación de matrices cuadradas de orden es $O(n^3)$. Pero los resultados nos daban un orden de $O(n^2.7)$ aproximadamente.

El algoritmo de Strassen de multiplicación de matrices consiste en multiplicar la matriz a trozos y nos promete un orden de $O(n^{log_2(7)+o(1)}) = O(log(7,2))$. Así que vamos a utilizar una regresión lineal múltiple con nuestros datos.

Regresión lineal múltiple

Os paso estos modelos

```r
lm_fit=lm(seconds~n+I(n^2)+I(n^3)+I(n^log(7,2)),data=data)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = seconds ~ n + I(n^2) + I(n^3) + I(n^log(7, 2)),
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11060 -0.12297 -0.00352  0.08026  2.63564
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.333e-01  5.131e-01   0.650   0.5194
## n              -1.769e-03  1.679e-03  -1.054   0.2977
## I(n^2)          3.135e-06  2.055e-06   1.525   0.1343
## I(n^3)          2.862e-09  1.310e-09   2.184   0.0343 *
## I(n^log(7, 2)) -1.671e-08  8.541e-09  -1.956   0.0568 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5532 on 44 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9916
## F-statistic:  1412 on 4 and 44 DF,  p-value: < 2.2e-16
```

```r
step(lm_fit)
```

```
## Start:  AIC=-53.29
## seconds ~ n + I(n^2) + I(n^3) + I(n^log(7, 2))
##
##                  Df Sum of Sq    RSS     AIC
## - n               1   0.33988 13.807 -54.066
## <none>                        13.467 -53.288
## - I(n^2)          1   0.71222 14.179 -52.762
## - I(n^log(7, 2))  1   1.17133 14.638 -51.201
## - I(n^3)          1   1.46005 14.927 -50.244
##
```

```
## Step:  AIC=-54.07
## seconds ~ I(n^2) + I(n^3) + I(n^log(7, 2))
##
##                   Df Sum of Sq    RSS     AIC
## <none>                         13.807 -54.066
## - I(n^2)           1    1.4322 15.239 -51.230
## - I(n^log(7, 2))   1    2.3190 16.126 -48.459
## - I(n^3)           1    3.1583 16.965 -45.972
```

```
##
## Call:
## lm(formula = seconds ~ I(n^2) + I(n^3) + I(n^log(7, 2)), data = data)
##
## Coefficients:
##    (Intercept)          I(n^2)          I(n^3)  I(n^log(7, 2))
##     -1.535e-01       1.028e-06       1.582e-09      -8.286e-09
```

```r
cor(data$n^3,data$n^log(7,2))
```

```
## [1] 0.9996034
```

```r
summary(update(lm_fit,~.-n-I(n^2)))
```

```
##
## Call:
## lm(formula = seconds ~ I(n^3) + I(n^log(7, 2)), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14618 -0.16832  0.01466  0.17051  2.91591
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.138e-01  1.508e-01   1.418    0.163
## I(n^3)           5.313e-10  8.434e-11   6.299 1.03e-07 ***
## I(n^log(7, 2)) -1.836e-09  4.297e-10  -4.272 9.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5756 on 46 degrees of freedom
## Multiple R-squared:  0.9913, Adjusted R-squared:  0.9909
## F-statistic:  2606 on 2 and 46 DF,  p-value: < 2.2e-16
```

```r
summary(lm(seconds~I(n^3),data=data))
```

```
##
## Call:
## lm(formula = seconds ~ I(n^3), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3557 -0.3527  0.0061  0.2295  3.5420
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.262e-01  1.288e-01  -1.756   0.0856 .
## I(n^3)       1.712e-10  2.777e-12  61.635   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6729 on 47 degrees of freedom
## Multiple R-squared:  0.9878, Adjusted R-squared:  0.9875
## F-statistic:  3799 on 1 and 47 DF,  p-value: < 2.2e-16
```

```
summary(update(lm_fit,~.-n-I(n^2)-I(n^log(7,2))))
```

```
##
## Call:
## lm(formula = seconds ~ I(n^3), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3557 -0.3527  0.0061  0.2295  3.5420
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.262e-01  1.288e-01  -1.756   0.0856 .
## I(n^3)       1.712e-10  2.777e-12  61.635   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6729 on 47 degrees of freedom
## Multiple R-squared:  0.9878, Adjusted R-squared:  0.9875
## F-statistic:  3799 on 1 and 47 DF,  p-value: < 2.2e-16
```

```
summary(update(lm_fit,~.-n-I(n^2)-I(n^3)))
```

```
##
## Call:
## lm(formula = seconds ~ I(n^log(7, 2)), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4387 -0.4378 -0.1328  0.4015  3.8727
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.218e-01  1.514e-01  -2.787  0.00765 **
## I(n^log(7, 2))  8.703e-10  1.634e-11  53.262  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7771 on 47 degrees of freedom
## Multiple R-squared:  0.9837, Adjusted R-squared:  0.9834
## F-statistic:  2837 on 1 and 47 DF,  p-value: < 2.2e-16
```

Cuestiones

Con estos modelos y algunos otros se pide:

1. Para el primer modelo. Comprobad que la suma de los residuos es 0. Calcular la desviación estándar de los residuos. Interpretar los resultados.

2. Para el primer modelo. Calcular los intervalos de confianza al 90% para los coeficientes. Con estos intervalos, los t-test y el ANOVA de la regresión discutir qué parámetros entran en el modelo.
3. Para el primer modelo. Dibujar los gráficos de los residuos contra los valores ajustados. Estudiar la normalidad de los residuos con un gráfico cuantil-cuantil
4. Utilizad el resto de modelos para decidir mediante la función step qué modelo más simple es el que se ajusta mejor a los datos.
5. Discutid qué modelo o modelos son los mejores para los datos y si es necesario ampliar la muestra de datos.