Travaux Pratiques - Modèl de Régression régularisée

Bealy MECH

10/4/2021

I. Test of significativity and model selection

a) Analyze and study the following instructions. Specify the underlaying theoretical model.

```
n=100;
X=cbind(((1:n)/n)^3, ((1:n)/n)^3);
Y=X%*%c(1,1)+rnorm(n)/4;
res=summary(lm(Y~X));
print(res);
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4777 -0.1544 -0.0197  0.1366  0.7550
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03171    0.03314  -0.957    0.341
## X1           2.03120    0.08617  23.573   <2e-16 ***
## X2                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.248 on 98 degrees of freedom
## Multiple R-squared:  0.8501, Adjusted R-squared:  0.8486
## F-statistic: 555.7 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
print(res$coef[2,4]);
```

```
## [1] 3.6042e-42
```

Compare the results provided by a multiple regression model and the results computed independently using two simple models. Conclusion.

```
reg1=lm(Y~X[,1]);
print(summary(reg1));
```

```
##
## Call:
## lm(formula = Y ~ X[, 1])
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -0.4777 -0.1544 -0.0197  0.1366  0.7550
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03171    0.03314  -0.957    0.341
## X[, 1]       2.03120    0.08617  23.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.248 on 98 degrees of freedom
## Multiple R-squared:  0.8501, Adjusted R-squared:  0.8486
## F-statistic: 555.7 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
reg2=lm(Y~X[,2]);
print(summary(reg2));
```

```
##
## Call:
## lm(formula = Y ~ X[, 2])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4777 -0.1544 -0.0197  0.1366  0.7550
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03171    0.03314  -0.957    0.341
## X[, 2]       2.03120    0.08617  23.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.248 on 98 degrees of freedom
## Multiple R-squared:  0.8501, Adjusted R-squared:  0.8486
## F-statistic: 555.7 on 1 and 98 DF,  p-value: < 2.2e-16
```

b) Execute the previous instructions several times (2 or 3 times) and describe the behaviour of the estimators of the coefficients. Compute the empirical correlation matrix. Instruction `cor()`.

```
cor(X[,1],X[,2])
```

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

II Model selection in a linear regression framework

```
tab = read.table(file="./UsCrime.txt", header=TRUE)
```