Lab 3

IMC 490: Machine Learning for IMC

4/12/2017

In this lab, we will be going over the following topics:
- drop1 vs anova
- interactions

drop1

For a full model with p predictors, `drop1()` trains p regression models, dropping one predictor with each model. Reports changes in MSS and RSS for each model, aiding in analysis of predictors.

Sum of Sq: Reduction in model sum of squares (MSS)
RSS: Value of residual sum of squares

In both anova and drop1, a point of confusion is that when RSS is reported, it is the ABSOLUTE value of RSS for each model, whereas when SumSq is reported, it is the CHANGE in SumSq for each predictor.

```
data(mtcars)
fit_full = lm(mpg ~ cyl + wt + hp, mtcars)
drop1(fit_full)
```

```
## Single term deletions
##
## Model:
## mpg ~ cyl + wt + hp
##        Df Sum of Sq    RSS    AIC
## <none>              176.62 62.665
## cyl     1    18.427 195.05 63.840
## wt      1   115.354 291.98 76.750
## hp      1    14.551 191.17 63.198
```

State the regression equation for each row in the drop1 output.

Let's manually verify these residuals.

1. Calculate TSS
2. Write code to verify the *RSS* value (291.98) for `wt` in the drop1 output
3. Write code to verify the *Sum of Sq* (115.35) for `wt` in the drop1 output
4. Verify the MSS, RSS, and TSS equality using the reduced model. Make sure you understand the intuition behind the equality statement.
5. Without looking at the model summary, can you guess which predictor is most significant? (add the parameter `test = "F"` to perform an F test.)

anova

`anova()` begins with an empty model and adds predictors in the order they are defined in the formula, evaluating sum of squares at each step.

```
anova(fit_full)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## cyl        1 817.71  817.71 129.6336 5.093e-12 ***
## wt         1 117.16  117.16  18.5740 0.0001822 ***
## hp         1  14.55   14.55   2.3069 0.1400152
## Residuals 28 176.62    6.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

State the regression equation for each row in the anova output.

Note that the *Sum Sq* reported is NOT the absolute sum of squares for each model, but rather the REDUCTION in sum of squares for each model.

1. Verify the *Sum Sq* for the second model.
2. Verify the TSS using the anova output.
3. Without looking at the model summary, can you guess which predictor is most significant?

Interactions

"An engineer suspects that the surface finish of metal parts is influenced by the type of paint used and the drying time. He selects three drying times and two types of paint." (from page 94 in course packet)

```
paint = data.frame(
  type = factor(c(rep(1,9), rep(2,9))),
  time = factor(rep(c(rep(20,3), rep(25,3), rep(30,3)), 2)),
  y = c(74,64,50, 73,61,44, 78,85,92, 92,86,68, 98,73,88, 66,45,85)
)
```

Read in the data and answer the following questions:

1. Fit a model to predict the quality of surface finish (`y`) using dry time (`time`), type of paint (`type`), and their interaction term.
2. Write the regression equation.
3. Create interaction plots of type and time. Explain in plain english the conclusion you draw from the plot.
4. Verify your conclusion with the model summary and drop1 output.

"An experiment was conducted to determine wheter either firing temperature of furnace position affects the baked density of a carbon anode." (from page 94 in the course packet)

```
anode = data.frame(
  pos = factor(c(rep(1,9), rep(2,9))),
  temp = factor(rep(c(rep(800,3), rep(825,3), rep(850,3)), 2)),
  density = c(570,565,583, 1063,1080,1043, 565,510,590, 528,547,521, 988,1026,1004, 526,538,532)
)
```

Read in the data and answer the following questions:

1. Fit a model to predict the density of the anode (`density`) against dry time (`time`), type of paint (`type`), and their interaction term.
2. Write the regression equation.
3. Create interaction plots of type and time. Explain in plain english the conclusion you draw from the plot.
4. Verify your conclusion with the model summary and drop1 output.