STAT021 Statistical Methods II

Lecture 20 MLR Model Building

Lu Chen
Swarthmore College
11/20/2018

Outline

- Model building strategies
  - Exhaustive modeling
  - Forward selection
  - Backward elimination
  - Stepwise procedure
- Examples
- Notes

Model building strategies

1. What are the potential predictors?

- Quantitative predictors
  - Linear terms (the quantitative predictors in the original scale)
  - Polynomial terms (2^{nd}, 3^{rd} order)?
  - Categorization?
  - Transformation based on a certain function (eg., $\log()$)
- Categorical predictors
  - Modeled as categorical or quantitative?
- Interaction terms (two-way, three-way?)
  - Usually we do not consider four-way or more complex interaction terms.

Model building strategies

2. Choose predictors from the potential predictors to build models

- **Exhaustive modeling**
  - Run all possible models with different combinations of the potential predictors.
- **Forward selection**
  - Start from a model without any predictors; keep adding predictors to the model based on some pre-defined criteria.
- **Backward elimination**
  - Start from a model with all potential predictors; keep deleting predictors from the model based on some pre-defined criteria.
- **Stepwise procedure**
  - Start from a model without any predictors; keep adding and deleting predictors from the model based on some pre-defined criteria.

Example

```r
# Input data
perch <- read.table("Perch.txt",sep="\t",header=T)
head(perch)
```

```
##   Obs Weight Length Width
## 1 104    5.9    8.8   1.4
## 2 105   32.0   14.7   2.0
## 3 106   40.0   16.0   2.4
## 4 107   51.5   17.2   2.6
## 5 108   70.0   18.5   2.9
## 6 109  100.0   19.2   3.3
```

- Potential predictors: $Length$, $Width$, $Length^2$, $Width^2$, $Length \times Width$

Exhaustive modeling - Example

All possible models with different combinations of the potential predictors:

1. $Weight \sim L$
2. $Weight \sim L + L^2$
3. $Weight \sim W$
4. $Weight \sim W + W^2$
5. $Weight \sim L + W$
6. $Weight \sim L + L^2 + W$
7. $Weight \sim L + W + W^2$
8. $Weight \sim L + L^2 + W + W^2$
9. $Weight \sim L + W + LW$
10. $Weight \sim L + L^2 + W + LW$
11. $Weight \sim L + W + W^2 + LW$
12. $Weight \sim L + L^2 + W + W^2 + LW$

Forward selection

1. Start from a model with no predictors and find the best single predictor (e.g. most significant P value based on t test).
2. Add each of the remaining predictors to the model separately, run the regression and find their individual P values:
   - If all of the P values are large (say, greater than 0.05), stop. The previous model is the best fitting model.
   - If any of the P values is small (less than 0.05), add the predictor with the smallest P value (that improves the model most significantly) to the model. Return to the start of Step 2.

- Note: The P value criterion above can be replaced by R^2_{adj}, Mallow's C_p or AIC.

Forward selection - Example

```r
# Model with no predictors
summary(m0 <- lm(Weight ~ 1, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   382.24      46.45   8.229  3.7e-11 ***
```

```r
# Add each predictor
summary(m1 <- lm(Weight ~ Length, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -652.787     43.407  -15.04   <2e-16 ***
## Length        35.001      1.398   25.03   <2e-16 ***
```

```r
summary(m2 <- lm(Weight ~ Width, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -509.289     35.594  -14.31   <2e-16 ***
## Width        188.115      7.038   26.73   <2e-16 ***
```

- $P_{Width} < P_{Length}$ because $t_{Width} > t_{Length}$. Add $Width$ to the model.

Forward selection - Example

```r
# Model with Width and one additional predictor
summary(m3 <- lm(Weight ~ Width + Length, data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -578.758     43.667 -13.254  < 2e-16 ***
## Width        113.500     30.265   3.750 0.000439 ***
## Length        14.307      5.659   2.528 0.014475 *
```

```r
summary(m4 <- lm(Weight ~ Width + I(Width^2), data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -31.735     91.908  -0.345    0.731
## Width        -25.635     39.487  -0.649    0.519
## I(Width^2)    20.934      3.827   5.470 1.24e-06 ***
```

- $P_{Width^2} < P_{Length}$. Add $Width^2$ to the model.

Forward selection - Example

```r
# Model with Width, Width^2 and one additional predictor
summary(m5 <- lm(Weight ~ Width + I(Width^2) + Length, data=perch))
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    36.348     59.832   0.607    0.546
## Width        -271.668     38.130  -7.125 3.13e-09 ***
## I(Width^2)     30.128      2.688  11.210 1.74e-15 ***
## Length         29.176      3.363   8.674 1.11e-11 ***
```

- $P_{Length} < 0.05$. Add *Length* to the model.

Forward selection - Example

```r
# Model with Width, Width^2, Length and one additional predictor
summary(m6 <- lm(Weight ~ Width + I(Width^2) + Length + I(Length^2), data=perch))
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    138.0015    60.4435   2.283 0.026621 *
## Width          -31.0365    73.9416  -0.420 0.676436
## I(Width^2)      10.0718     5.9717   1.687 0.097793 .
## Length         -15.2436    12.4688  -1.223 0.227124
## I(Length^2)      0.6065     0.1652   3.672 0.000577 ***
```

```r
summary(m7 <- lm(Weight ~ Width * Length + I(Width^2), data=perch))
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     114.5331    60.3972   1.896  0.06359 .
## Width           -91.2681    66.6836  -1.369  0.17710
## Length           -4.0403    10.8834  -0.371  0.71200
## I(Width^2)       -0.5327     9.9434  -0.054  0.95748
## Width:Length      5.3281     1.6734   3.184  0.00248 **
```

- $P_{Length^2} < P_{Length \times Width}$. Add $Length^2$ to the model.

Forward selection - Example

```r
# Model with Width, Width^2, Length, Length^2 and the interaction: full model
summary(m8 <- lm(Weight ~ Width * Length + I(Width^2) + I(Length^2), data=perch))
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    156.3486    61.4152   2.546   0.0140 *
## Width           20.9772    82.5877   0.254   0.8005
## Length         -25.0007    14.2729  -1.752   0.0860 .
## I(Width^2)      34.4058    18.7455   1.835   0.0724 .
## I(Length^2)      1.5719     0.7244   2.170   0.0348 *
## Width:Length    -9.7763     7.1455  -1.368   0.1774
```

- $P_{Length \times Width} > 0.05$. Stop at the previous model.
- Forward selection selects the model with $Width$, $Width^2$, $Length$ and $Length^2$ as the final model.

Forward selection - Notes

- Forward selection evaluates less models than exhuastive modeling.
- Some models will never be evaluated.
- It is based on only one criterion.
- Once a predictor is added to the model, this stragety will never removes the predictor even if it becomes insignificant.

Backward elimination

1. Start by fitting the full model (the model that includes all potential predictors).
2. Identify the predictor for which the individual t test produces the largest P value:
   - If that P value is large (say, greater than 0.05), eliminate the predictor to produce a smaller model. Fit the model and return to the start of Step 2.
   - If the P value is small (less than 0.05), stop since all of the predictors in the model are "significant."

- Note: The P value criterion above can be replaced by R^2_{adj}, Mallow's C_p or AIC.

Backward elimination - Example

```r
## Full model
summary(m1 <- lm(Weight ~ Width * Length + I(Width^2) + I(Length^2), data=perch))
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      156.3486    61.4152   2.546   0.0140 *
## Width             20.9772    82.5877   0.254   0.8005
## Length           -25.0007    14.2729  -1.752   0.0860 .
## I(Width^2)        34.4058    18.7455   1.835   0.0724 .
## I(Length^2)        1.5719     0.7244   2.170   0.0348 *
## Width:Length      -9.7763     7.1455  -1.368   0.1774
```

- $Width$ is not significant and has the largest P-value, remove it.
- Then $Width^2$ and $Length \times Width$ must be removed.

Backward elimination - Example

```r
## Remove the predictor that is not and least significant
summary(m2 <- lm(Weight ~ Length + I(Length^2), data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 128.34533   78.77870   1.629  0.10920
## Length      -21.02388    5.41770  -3.881  0.00029 ***
## I(Length^2)   0.90862    0.08689  10.458 1.72e-14 ***
```

- Both terms are significant. Stop.
- Backward elimination selects the model with $Length$ and $Length^2$ as the final model.

Notes

- It only evaluates 2 models. Many models were not evaluated.
- It is based on only one criterion.
- It may accidentally remove predictor(s) that is important to the model.

Stepwise procedure

A stepwise procedure starts from forward selection, but after any new predictor is added to the model, it uses backward elimination to delete any predictors that have become redundant in the model.

- Forward selection may result in redundant predictor(s) in the model.
- Backward elimination may accidentally delete predictor(s) that is important to the model.
- The stepwise procedure may achieve a balance between the two.

Stepwise procedure - Example

```r
# Model with no predictors
summary(m0 <- lm(Weight ~ 1, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   382.24      46.45   8.229  3.7e-11 ***
```

```r
# Add each predictor
summary(m1 <- lm(Weight ~ Length, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -652.787     43.407  -15.04   <2e-16 ***
## Length        35.001      1.398   25.03   <2e-16 ***
```

```r
summary(m2 <- lm(Weight ~ Width, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -509.289     35.594  -14.31   <2e-16 ***
## Width        188.115      7.038   26.73   <2e-16 ***
```

- $P_{Width} < P_{Length}$ because $t_{Width} > t_{Length}$. Add *Width* to the model.
- Look backward: any predictor is insignificant in the current model?
- *Width* is significant, keep it.

Stepwise procedure - Example

```r
# Add one additional predictor
summary(m3 <- lm(Weight ~ Width + I(Width^2), data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -31.735     91.908  -0.345    0.731
## Width        -25.635     39.487  -0.649    0.519
## I(Width^2)    20.934      3.827   5.470 1.24e-06 ***
```

```r
summary(m4 <- lm(Weight ~ Width + Length, data=perch))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -578.758     43.667 -13.254  < 2e-16 ***
## Width        113.500     30.265   3.750 0.000439 ***
## Length        14.307      5.659   2.528 0.014475 *
```

- $P_{Width^2} < P_{Length}$. Add $Width^2$ to the model.
- Look backward: any predictor is insignificant in the current model?
- $Width$ becomes insignificant, we keep it for the quadratic term is in the model.

Stepwise procedure - Example

```r
# Add one additional predictor
summary(m5 <- lm(Weight ~ Width + I(Width^2) + Length, data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     36.348     59.832   0.607    0.546
## Width         -271.668     38.130  -7.125 3.13e-09 ***
## I(Width^2)      30.128      2.688  11.210 1.74e-15 ***
## Length          29.176      3.363   8.674 1.11e-11 ***
```

- $P_{Length} < 0.05$. Add *Length* to the model.
- Look backward: any predictor is insignificant in the current model?
- All predictors are significant. Keep them all.

Stepwise procedure - Example

```r
# Add one additional predictor
summary(m6 <- lm(Weight ~ Width + I(Width^2) + Length + I(Length^2), data=perch))
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 138.0015    60.4435   2.283 0.026621 *
## Width       -31.0365    73.9416  -0.420 0.676436
## I(Width^2)   10.0718     5.9717   1.687 0.097793 .
## Length      -15.2436    12.4688  -1.223 0.227124
## I(Length^2)   0.6065     0.1652   3.672 0.000577 ***
```

```r
summary(m7 <- lm(Weight ~ Width * Length + I(Width^2), data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   114.5331    60.3972   1.896  0.06359 .
## Width         -91.2681    66.6836  -1.369  0.17710
## Length         -4.0403    10.8834  -0.371  0.71200
## I(Width^2)     -0.5327     9.9434  -0.054  0.95748
## Width:Length    5.3281     1.6734   3.184  0.00248 **
```

- $P_{Length^2} < P_{Length \times Width}$. Add $Length^2$ to the model.

Stepwise procedure - Example

```r
summary(m6 <- lm(Weight ~ Width + I(Width^2) + Length + I(Length^2), data=perch))
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 138.0015    60.4435   2.283 0.026621 *
## Width       -31.0365    73.9416  -0.420 0.676436
## I(Width^2)   10.0718     5.9717   1.687 0.097793 .
## Length      -15.2436    12.4688  -1.223 0.227124
## I(Length^2)   0.6065     0.1652   3.672 0.000577 ***
```

- Look backward: any predictor is insignificant in the current model?
- *Width* is insigificant. Remove it and the quadratic term.

```r
summary(m8 <- lm(Weight ~ Length + I(Length^2), data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 128.34533   78.77870   1.629  0.10920
## Length      -21.02388    5.41770  -3.881  0.00029 ***
## I(Length^2)   0.90862    0.08689  10.458 1.72e-14 ***
```

- All predictors are significant. Keep them all.

Stepwise procedure - Example

```
# Add one additional predictor
summary(m8 <- lm(Weight ~ Length + I(Length^2) + Width, data=perch))
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 147.12090   61.25958   2.402   0.0199 *
## Length      -34.71805    4.78840  -7.250 1.97e-09 ***
## I(Length^2)   0.86134    0.06794  12.679  < 2e-16 ***
## Width        91.09772   15.20858   5.990 2.00e-07 ***
```

- *Width* is significant. Keep it.
- Look backward: any predictor is insignificant in the current model?
- All predictors are significant.

Stepwise procedure - Example

```r
# Add one additional predictor
summary(m9 <- lm(Weight ~ Length + I(Length^2) + Width + I(Width^2), data=perch))
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 138.0015    60.4435   2.283 0.026621 *
## Length      -15.2436    12.4688  -1.223 0.227124
## I(Length^2)   0.6065     0.1652   3.672 0.000577 ***
## Width       -31.0365    73.9416  -0.420 0.676436
## I(Width^2)   10.0718     5.9717   1.687 0.097793 .
```

```r
summary(m10 <- lm(Weight ~ Length * Width + I(Length^2), data=perch))
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    136.1045    61.8038   2.202   0.0322 *
## Length         -19.2196    14.2406  -1.350   0.1831
## Width           -3.5967    83.3663  -0.043   0.9658
## I(Length^2)      0.4300     0.3795   1.133   0.2625
## Length:Width     2.6672     2.3089   1.155   0.2534
```

- Both $Width^2$ and $Length \times Width$ are not significant. Stop.

- Step precedure selects the model with $Length$, $Width$ and $Length^2$ as the final model.

Three strategies

Forward selection

$$Weight \sim L + L^2 + W + W^2$$

Backward elimination

$$Weight \sim L + L^2$$

Stepwise procedure

$$Weight \sim L + W + L^2$$

- The three strategies results in three different models. There are always some models that cannot be evaluated.
- Therefore, we should take the models found by these strategies as **potential** models and apply other criteria to further select the best model.

Caution about model building

- In R, the `step()` function automatically selects models for you using these three strategies based on **AIC** rather than t test. Note: the AIC value from the `step()` function is computed differently from the `AIC()` function.
- However, do not rely on automated techniques. Even if the R functions for forward, backward and stepwise modeling give us the same model, it does not mean that it is the THE model.
- It is always the responsibility of the modeler to
  - Think about the possible models
  - Conduct diagnostic procedures (checking error assumptions and searching for unusual points)
  - Only use models that make sense
- Criticisms of stepwise procedures: multiple comparisons and thus inflated type I error rate.

Model building using `step()`

```r
# Model without any predictors
zeroModel <- lm(Weight ~ 1, data=perch)

# Full model
fullModel <- lm(Weight ~ Length*Width + I(Length^2) + I(Width^2), data=perch)

# Forward selection
forward <- step(zeroModel, scope=list(upper = fullModel), direction="forward")

# Backward elimination
backward <- step(fullModel, scope=list(lower = zeroModel), direction="backward")

# Stepwise procedure
stepwise <- step(zeroModel, scope=list(upper = fullModel), direction="both")
```

- Note: *full model* could be different based on your pool of potential predictors.
- For more examples and usage of the `step()` function, see *Lecture20_Examples.Rmd.*

Some notes

- If any interaction term is present, always include the main effect terms.
- If any polynomial term is present, always include the lower-order terms.
- The `step()` function does not follow the rules above. Add the main effect and lower-order terms if the final model should have them.
- Model comparison should be based on several statistics (R^2_{adj}, C_p, AIC and/or nested F test) rather than one single statistic.
- Two or more models may have very similar model fitting. In general, we choose the simpler one (with less/cheaper predictors or easier to understand).
- Whether to model a categorical predictor as a categorical or quantitative variable depends on its practical meaning and/or model fitting.
- Whether to categorize a quantitative predictor depends on its practical meaning and/or model fitting.
- Check error assumptions and unusual points for the final model and possibly one or several other models in the process of model building.