MATH411 | Fall 2018 | Exam III

Paul Tomosky

Friday, 12/14/2018

Icing the Kicker

Before a placekicker attempts a field goal in a pressure situation, an opposing team may call a time-out. The purpose of this time-out is to give the kicker more time to think about the attempt in the hopes that this extra time will cause him to become more nervous and lower his probability of success. This strategy is often referred to as "**icing the kicker**." `Icing the Kicker.csv` contains the data from the 2002 and 2003 National Football League seasons to investigate whether or not the strategy actually lowers the probability of success when implemented. Below are descriptions of the variables available in this file:

- `GameNum`: Identifies the year and game
- `Kicker`: Last name of kicker
- `Good`: Response variable ("Y" = success, "N" = failure)
- `Distance`: Length in yards of the field goal
- `Weather`: Levels of "Clouds", "Inside", "SnowRain", and "Sun"
- `Wind15`: 1 if wind speed is $\geq$ 15 miles per hour and the placekick is outdoors, 0 otherwise.
- `Temperature`: Levels of "Nice" (40F < temperature < 80F or inside a dome), "Cold" (temperature $\leq$ 40F and outdoors), and "Hot" (temperature $\geq$ 80F and outdoors)
- `Grass`: 1 if kicking on a grass field, 0 otherwise
- `Pressure`: "Y" if attempt is in the last 3 minutes of a game and a successful field goal causes a lead change, "N" otherwise
- `Ice`: 1 if Pressure = 1 and a time-out is called prior to the attempt, 0 otherwise

Your goal in this exam is to build the best model for predicting `Good` by using `Distance`, `Weather`, `Wind15`, `Temperature`, `Grass`, `Pressure`, and `Ice` as potential predictors. To complete this goal,

1. Let's use the **overall acuracy** as the criterion for measuring your model, i.e., your goal is to find a model that will maximize the overall accuracy.

2. In order to calculate the **overall acuracy** value, you need first to split the data into **train set** and **test set**. Let's do so by using 80% of the data as the **train set** and the remaining 20% as the **test set** (When you are splitting the data, please use

`set.seed(2018667)`). Then the **overall acuracy** can be calculated based on the **test set**.

3. To submit:

3.1 You only need to provide your final model, i.e., your best model along with the overall **overall acuracy**. **But you need to describe your strategy of how you achieve this model**.

3.2 Save and submit the **test set** as a `csv` file. Within the file add a new column that contains your predicts. Also provide a **confusion matrix** for your predictions.

Solutions:

Final Model:

```
mod1 = glm(success ~ 1 + distance + weather + iced + wind15:weather + grass:distance +
           pressure:grass + iced:distance,
           family = binomial,
           data   = train)
```

Confusion Matrix:

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0  61  82
         1  22 236

               Accuracy : 0.7406
                 95% CI : (0.6948, 0.7829)
    No Information Rate : 0.793
    P-Value [Acc > NIR] : 0.9951

                  Kappa : 0.3765
 Mcnemar's Test P-Value : 7.233e-09

            Sensitivity : 0.7421
            Specificity : 0.7349
         Pos Pred Value : 0.9147
         Neg Pred Value : 0.4266
              Precision : 0.9147
                 Recall : 0.7421
                     F1 : 0.8194
             Prevalence : 0.7930
         Detection Rate : 0.5885
   Detection Prevalence : 0.6434
      Balanced Accuracy : 0.7385

       'Positive' Class : 1
```

Strategy:

- Load and format the data
- Split into test / train
- Create the large model
- Use glmulti to find models with interactions
- Compare the best models
- Select best model based on
    - AIC
    - Internal estimate of accuracy
    - Cross-validation estimate of accuracy
- Use best model for prediction
- Create the confusion matrix
- Save and output the data