Lab #6 - Predictive Regression II

Econ 224

September 11th, 2018

Predicting College Football Games

The data for duplicating the football results in Table 10-1 on page 165 and in Table 10-3 on page 170 are in the file: football.txt. For example, for the regression in Table 10-3 have the students regress SPREAD on LV, H, SAG, BIL, COL, DUN, and REC (no constant term) using the 1,582 observations.

Predictors: 6 ranking systems, win-loss record, and home team variable.

1. Matthews/Scripps Howard (MAT)
2. Jeff Sagarin's *USA Today* (SAG)
3. Richard Billingsley (BIL)
4. *Atlanta Journal-Constitution* Colley Matrix (COL)
5. Kenneth Massey (MAS)
6. Dunkel (DUN)
7. System using *only* won-loss records (REC)

Data for 1998, 1999, 2000, 2001. Ten weeks of data for each year, beginning with week 6 for a total of 1,582 games. Division I-A teams: 117 in 2001, 115 in 2000, 144 in 1999, and 112 in 1998. Data courtesy of Professor Ray Fair of Yale University.

This example is base on Chapter 10 of Ray Fair's book *Predicting Presidential Elections and Other Things*

Variable we are predicting is *point spread* in a game (SPREAD). Two teams in a game: A and B. The point spread is team A's score minus team B's score. For example, if A beat B 28 to 13, SPREAD equals 15. If B beats A 28 to 13, SPREAD equals -15.

How are the predictors constructed? Difference in rankings for team A minus team B in the week when the game is scheduled to take place. For example if Richard Billingsley ranks Stanford #10 and UCLA #22, the predictor BIL equals 11 if we treat Stanford as team A and -11 if we treat UCLA as team A. This is how MAT, SAG, BIL, COL, MAS, and DUN are constructed. The remaining variable REC is different since it is based on *win-loss records*

Second Part: KNN, out-of-sample, etc

- Implement KNN in one dimension and test on some simulated data
- predicting birthweight based on gestation

Child Health and Development Studies conducted at the Oakland, CA, Kaiser Foundation Hospital. The variables are

```
bwt: baby's weight in ounces at birth
gestation: duration of pregnancy in days
parity: parity indicator (first born = 1, later birth = 0)
age: mother's age in years
height: mother's height in inches
weight: mother's weight in pounds (during pregnancy)
smoke: indicator for whether mother smokes (1=yes, 0=no)
```

```
Bwt = read.csv("http://people.reed.edu/~jones/141/Bwt.dat")
```