Predicting Wins and Losses in NFL

Immanuel Williams

Data Mining Fall 2014

Data Derivation

- Data is based on the last 10 years of NFL data
- There are 32 teams and each team plays 16 games
- Only look at data from previous 2 years and last 6 games of the season
- Variables: yards, points, turnovers, point difference and average wins
- **Manipulation**: 1258 games 45 variables
- Split the data into two datasets Train (958 games) and Test (300 games)

Main Methods Used

- SVM, OLS, LR, LDA, QDA, SVM, KNN, Ridge and LASSO was use on the 45 variable model

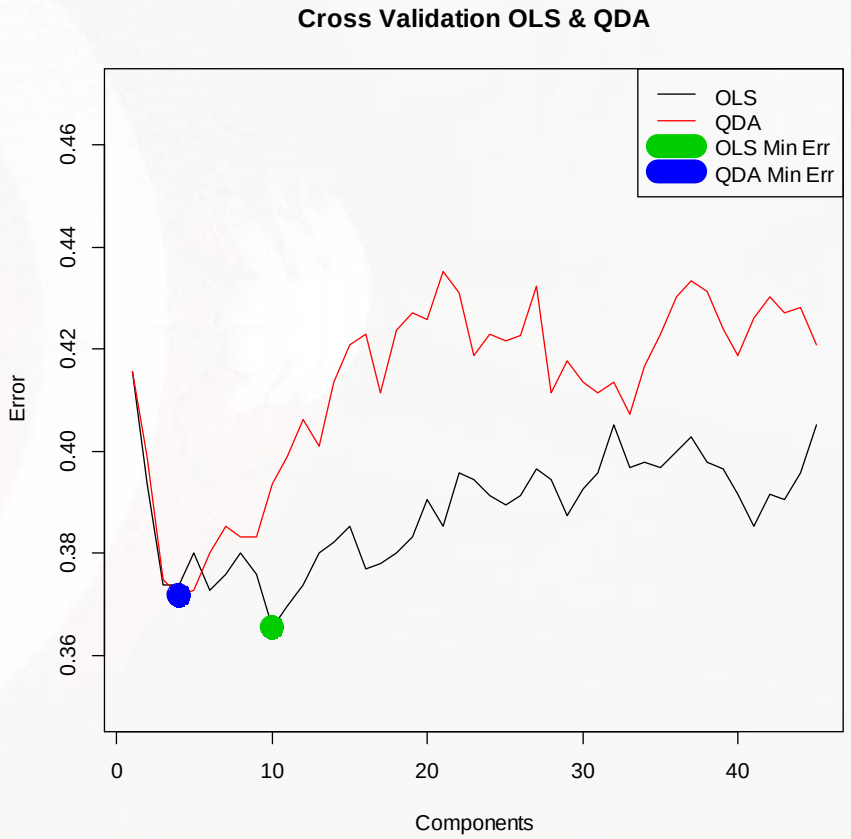
All 45 Variables			Conditions
	Training	Test	
OLS	0.338	0.390	
LR	0.338	0.475	
LDA	0.338	0.390	
QDA	0.247	0.430	
SVM	0.344	0.383	Best Cost = 0.01 & Best γ=0.022
KNN	0.211	0.370	N=3 & N=22
Ridge	0.338	0.373	λ=0.011
Lasso	0.336	0.373	λ=0.001

Main Methods Used

- Cross Validation QDA and OLS on the 45 variable model

Minimum OLS Error: 0.365
10 Principal Components

Minimum QDA Error: 0.371
4 Principal Components



Another Method

- Basis Expansion on current year variables→ # of increases to 180

All 180 Variables			Conditions
	Training	Test	
OLS	0.303	0.443	
LR	0.284	0.443	
LDA	0.303	0.443	
QDA	0.197	0.476	
SVM	0.338	0.353	Best Cost = 0.001 & Best γ=0.005
KNN	0.213	0.423	N=2 & N=21
Ridge	0.303	0.45	λ=0.0031
Lasso	0.302	0.403	λ=0.0051