How Do Injuries in the NFL Affect the Outcome of the Game

Andy Sun

STATS 50 2014

Motivation

- NFL injury surveillance shows upward trend in both the number and severity of injuries

- Packers won 2010 Super Bowl with 15 players on injured reserve

- What is the breaking point?

Two Approaches

- Differences in Probabilities of Winning; pre and post game

- Adjusted Games Loss

Differences in Winning Probability

- Let X = Whether or not the favored team won $\rightarrow$ 1 or 0

- Let p = predicted probability of the favored team winning

- Let Y = Diff. in Winning Prob. = X - p

- Predicted probability of winning is based on point **spread**, and NFL Betting Odds.
  - Ex/

Favored	Spread	Underdog
Broncos	-3.5	Chiefs

- Spread -> Probability
  - The more negative the spread, the larger the probability of winning

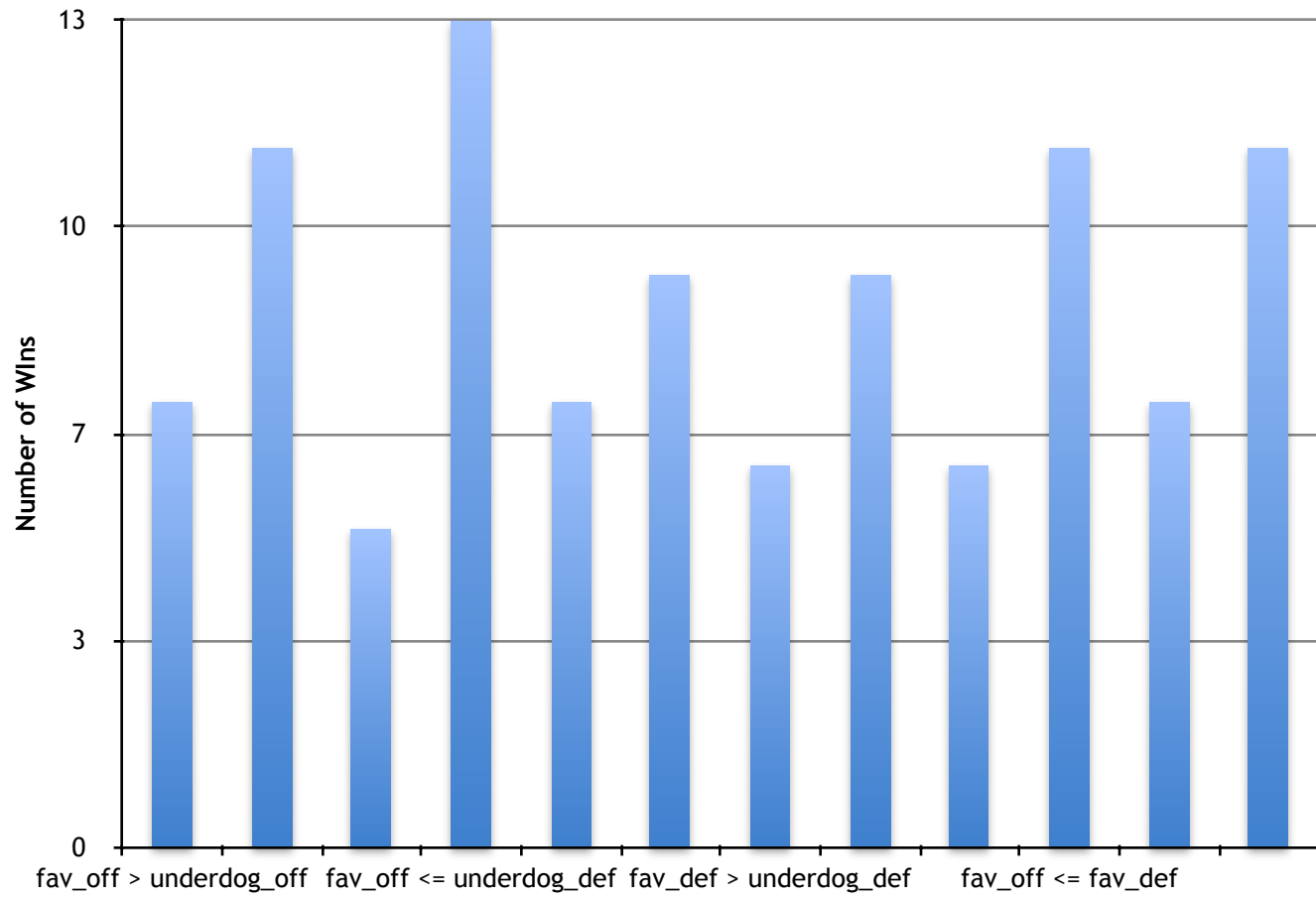| Spread	Probability of Winning
24	4%
0	50%
-24	96%

If the value of Y is…..

- Positive → favored team wins
  - The smaller the value of Y, the more heavily favored the team was to win

- Negative → favored team loses
  - The more negative the value of Y, the bigger the upset

Bins based on Injuries Sustained

- Look at the difference in winning probability based on the number of injuries for that week

- Ex/ Fav_off > Udog_def

- Ex/ Fav_total > Udog_total

Take the Averages in Each Bin

Bin	Avg Diff in Winning Probability
fav_off > underdog_off	-0.117
fav_off <= underdog_off	0.594
fav_off > underdog_def	0.12
fav_off <= underdog_def	-0.079
fav_def > underdog_off	0.025
fav_def <= underdog_off	0.0195
fav_def > underdog_def	0.1197
fav_def <= underdog_def	-0.186
fav_off > fav_def	-0.1346
fav_off <= fav_def	0.05
fav_total > underdog_total	0.0676
fav_total <= underdog_total	-0.075

Favored Team Loses

Bin	Avg Diff in Winning Probability
fav_off > underdog_off	-0.117
fav_off <= underdog_def	-0.079
fav_def <= underdog_def	-0.186
fav_off > fav_def	-0.1346
fav_total <= underdog_total	-0.075

✓ Favored Offense has more injuries than opposing offense

? Favored team has less injuries overall than the underdog team

Favored Team Wins

Bin	Avg Diff in Winning Probability
fav_off <= underdog_off	0.594
fav_off > underdog_def	0.12
fav_def > underdog_off	0.025
fav_def <= underdog_off	0.0195
fav_def > underdog_def	0.1197
fav_off <= fav_def	0.05
fav_total > underdog_total	0.0676

✓ Favored defense has injuries than opposing offense

? Favored team has more injuries overall than the underdog team

Challenges

- Data Collection
  - No readily usable data aggregated together
  - Lots of preprocessing needed

- Sample Size
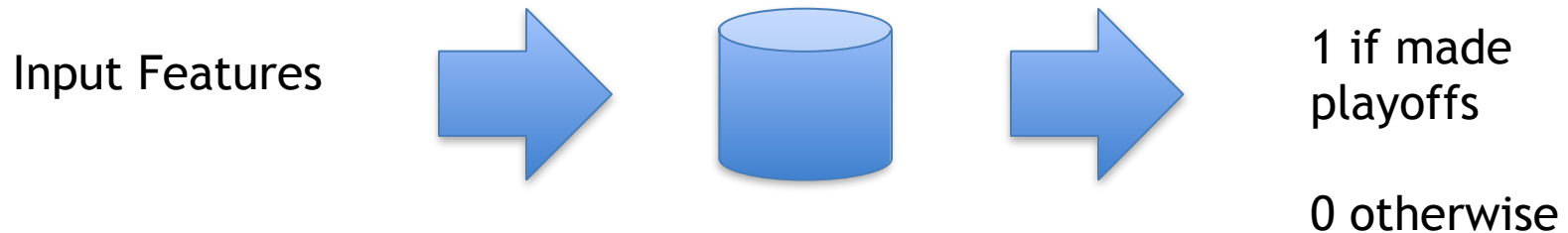 - Small sample size due to challenges in collecting data

Adjusted Games Loss (AGL)

- Attempt to find more conclusive results!

- Quantifies how much teams are affected by injuries to starters and injury replacements, and by player listing status on NFL Injury Report.

- http://www.footballoutsiders.com/stat-analysis/2014/2013-adjusted-games-lost

"Brute Force"

- With difference in winning probability, very careful selection of data

- With AGL, took more of a brute force approach with **machine learning**

- 2013 – 2014 NFL Season

- Use x inputs to "predict" whether or not a certain team made it to the playoffs

Input Features → [database] → 1 if made playoffs
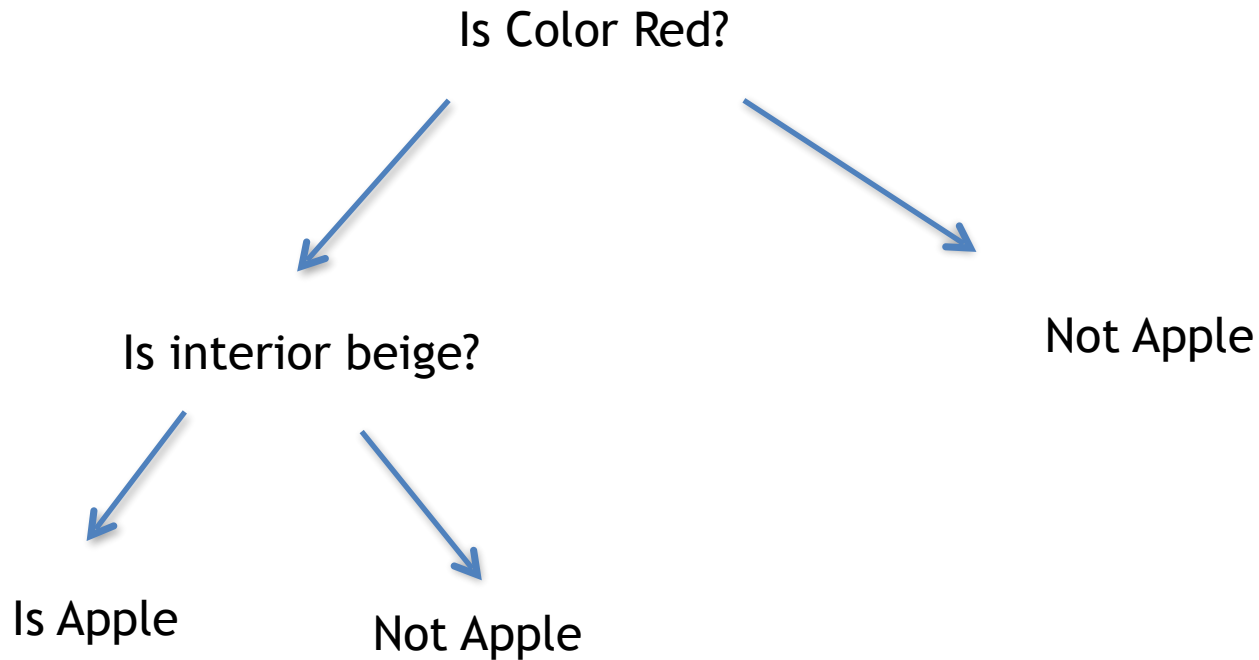
0 otherwise

- Analyze the most important features used in prediction

Binary Classification Tree

- Analogy → Fruits

- Going to classify whether an input is an apple or not

Features	Label
Red ext, beige int, medium sized, low	*Apple*
Red ext, purple int, small sized, med	*Grape*
Red ext, gray int, small sized, lost of	*Pomegranate*
Green ext, red int, large sized, lots of	*Watermelon*

Is Color Red?

Is interior beige?

Not Apple
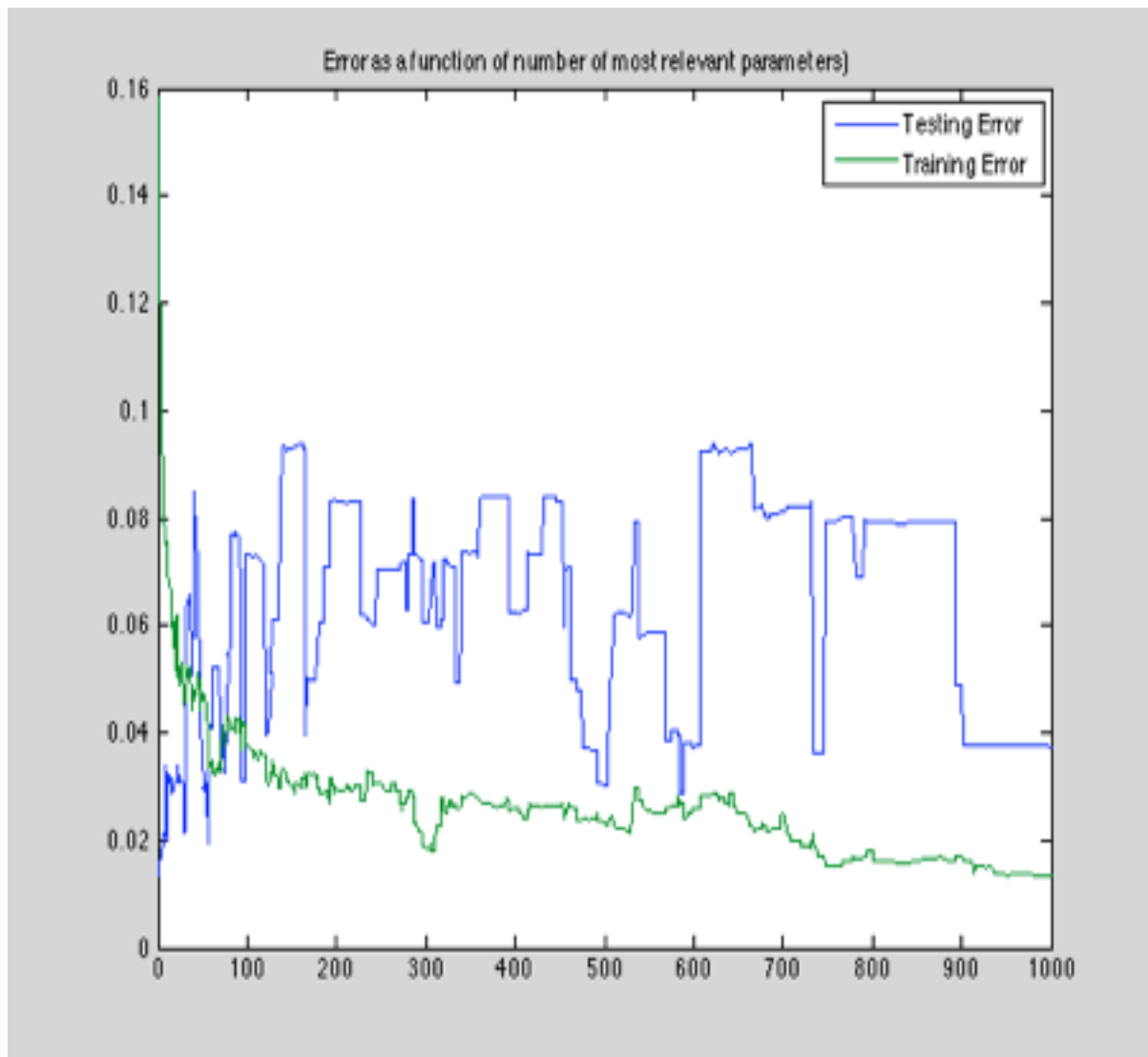
Is Apple

Not Apple

Notice how we never have to ask about size or the
Number of seeds.

Back to football

- Create feature vector for each NFL team

- Create corresponding label indicating whether or not that team made the play offs

- Feature Selection: How important of a feature is AGL? How high up in the tree structure is it?

The Predictive Model

Error as a function of number of most relevant parameters)

How important is AGL as a Feature?

- Not very important

- This may also me due to human error in deciding which features to use

- Chose this approach so that I could look at lots of data, but same problem as before: not enough quickly usable injury databases

Inconclusive

- Need to do gather and format more data

Thank You!