Determining the Outcome of NBA Games

Frank Longueira & Matthew Smarsch
Professor Keene
Fall 2015

Introduction

- Classifying the outcome of sporting events has obvious applications (i.e. fantasy sports, betting, etc.)

- Basketball lends itself nicely to analysis due to a large number of games played and emphasis on team play

Previous Work

- "Prediction of NBA games based on Machine Learning Methods"; Renato Amorim Torres; University of Wisconsin Madison

- "Predicting the Betting Line in NBA Games"; Chang, Dade, Lipman, Mills; Stanford University

Time to walk through our solution...

Will there be an Autoencoder...

Stay Tuned!

Data Collection and Processing

- Used Beautiful Soup's Python API to get and parse HTML for necessary pages

- Collected various stats
  - PTs, STLs, REBs, ASTs, TOs, BLKs, GP

- Prepared test set by calculating a running average of all statistics through each game of the season

Data Source

G	Date					Opponent			Tm	Opp	W	L	Streak	Notes
1	Wed, Oct 29, 2014	8:00p EST		Box Score		Chicago Bulls	L		80	104	0	1	L 1	
2	Thu, Oct 30, 2014	8:00p EST		Box Score	@	Cleveland Cavaliers	W		95	90	1	1	W 1	
3	Sun, Nov 2, 2014	7:30p EST		Box Score		Charlotte Hornets	W		96	93	2	1	W 2	
4	Tue, Nov 4, 2014	7:30p EST		Box Score		Washington Wizards	L		83	98	2	2	L 1	
5	Wed, Nov 5, 2014	7:30p EST		Box Score	@	Detroit Pistons	L		95	98	2	3	L 2	
6	Fri, Nov 7, 2014	7:30p EST		Box Score		Brooklyn Nets	L		99	110	2	4	L 3	
7	Sat, Nov 8, 2014	7:30p EST		Box Score	@	Atlanta Hawks	L		96	103	2	5	L 4	
8	Mon, Nov 10, 2014	7:30p EST		Box Score		Atlanta Hawks	L		85	91	2	6	L 5	
9	Wed, Nov 12, 2014	7:30p EST		Box Score		Orlando Magic	L		95	97	2	7	L 6	
10	Fri, Nov 14, 2014	7:30p EST		Box Score		Utah Jazz	L		100	102	2	8	L 7	
11	Sun, Nov 16, 2014	1:00p EST		Box Score		Denver Nuggets	W		109	93	3	8	W 1	
12	Tue, Nov 18, 2014	8:00p EST		Box Score	@	Milwaukee Bucks	L		113	117	3	9	L 1	
13	Wed, Nov 19, 2014	8:00p EST		Box Score	@	Minnesota Timberwolves	L		99	115	3	10	L 2	
14	Sat, Nov 22, 2014	7:30p EST		Box Score		Philadelphia 76ers	W		91	83	4	10	W 1	
15	Mon, Nov 24, 2014	8:00p EST		Box Score	@	Houston Rockets	L		86	91	4	11	L 1	
16	Wed, Nov 26, 2014	7:30p EST		Box Score	@	Dallas Mavericks	L	OT	102	109	4	12	L 2	
17	Fri, Nov 28, 2014	8:00p EST		Box Score	@	Oklahoma City Thunder	L		78	105	4	13	L 3	
18	Sun, Nov 30, 2014	7:30p EST		Box Score		Miami Heat	L		79	86	4	14	L 4	
19	Tue, Dec 2, 2014	7:30p EST		Box Score		Brooklyn Nets	L		93	98	4	15	L 5	
20	Thu, Dec 4, 2014	8:00p EST		Box Score		Cleveland Cavaliers	L		87	90	4	16	L 6	
G	Date					Opponent			Tm	Opp	W	L	Streak	Notes
21	Fri, Dec 5, 2014	7:00p EST		Box Score	@	Charlotte Hornets	L		102	103	4	17	L 7	
22	Sun, Dec 7, 2014	7:30p EST		Box Score		Portland Trail Blazers	L		99	103	4	18	L 8	
23	Tue, Dec 9, 2014	8:00p EST		Box Score	@	New Orleans Pelicans	L		93	104	4	19	L 9	
24	Wed, Dec 10, 2014	8:30p EST		Box Score	@	San Antonio Spurs	L		95	109	4	20	L 10	
25	Fri, Dec 12, 2014	7:30p EST		Box Score		Boston Celtics	W		101	95	5	20	W 1	
26	Sun, Dec 14, 2014	7:30p EST		Box Score		Toronto Raptors	L	OT	90	95	5	21	L 1	
27	Tue, Dec 16, 2014	8:00p EST		Box Score		Dallas Mavericks	L		87	107	5	22	L 2	
28	Thu, Dec 18, 2014	8:00p EST		Box Score	@	Chicago Bulls	L		97	103	5	23	L 3	
29	Sat, Dec 20, 2014	1:00p EST		Box Score		Phoenix Suns	L		90	99	5	24	L 4	
30	Sun, Dec 21, 2014	3:30p EST		Box Score	@	Toronto Raptors	L		108	118	5	25	L 5	
31	Thu, Dec 25, 2014	12:00p EST		Box Score		Washington Wizards	L		91	102	5	26	L 6	
32	Sat, Dec 27, 2014	10:00p EST		Box Score	@	Sacramento Kings	L	OT	129	135	5	27	L 7	
33	Sun, Dec 28, 2014	9:00p EST		Box Score	@	Portland Trail Blazers	L		79	101	5	28	L 8	
34	Wed, Dec 31, 2014	6:00p EST		Box Score	@	Los Angeles Clippers	L		78	99	5	29	L 9	

Team Totals 240 37 96 .385 12 29 .414 20 26 .769 23 36 59 23 5 3 15 15 106

Feature Space

- Analyzed significant team stats (PPG, Win %, RPG, etc.)
  - Both offensive and defensive
  - Assumed these team stats are jointly Gaussian

- Take the relative difference of competing teams' stats
  - This is the primary measure that serves as our feature vector!
  - Our initial feature space was 25 dimensional

- In addition, we normalized our feature space using *zscore() in* MATLAB
  - The mean was subtracted from each feature and then divided by its standard deviation
  - This maintains statistical importance, while ensuring classifier accuracy

Training/Test Sets

- Training set is generated based on 2008–2013* season schedules and outcomes
  - Per season, we associate a feature vector involving the overall season stats per game for each team
  - Each label is: 0 if the away team loses and 1 if the away team wins

- A test set was generated using the 2014-15 season
  - For each game, the ML estimate of each team's stats was taken to be our feature vector
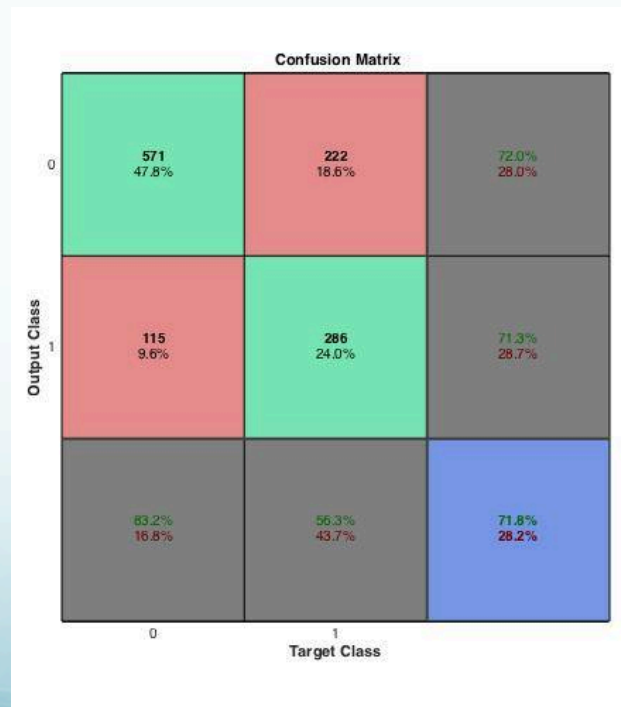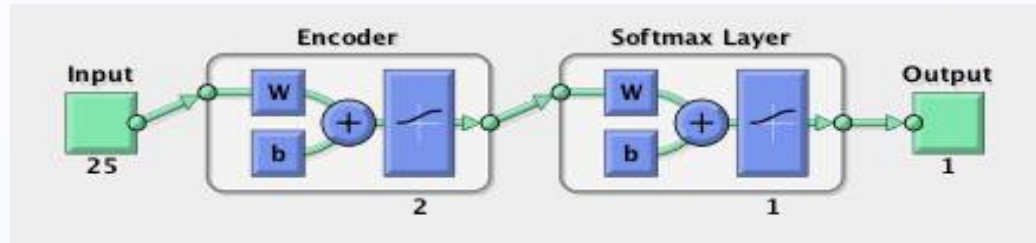
* Excluding lockout-shortened 2011 season

Classification Methods

- Linear Regression

- K-Nearest Neighbors

- SVM

- **<u>Autoencoder</u>**

Results

Classifier	Accuracy
Linear Regression	71.1%
K-Nearest Neighbors	69.3%
Support Vector Machine	71.0%
Autoencoder	71.8%

Autoencoder

Conclusions

- Based on our validation set, our model performs at or above benchmark metrics for each classifier

- The autoencoder implementation had the highest classification accuracy, but all models were within 1-2% accuracy

- Application can be readily extended to the playoffs as well

Questions?