NFL FANTASY FOOTBALL

Project 2: Predicting NFL Fantasy Points

Jan 26, 2018

Jaydon Youngho Jung

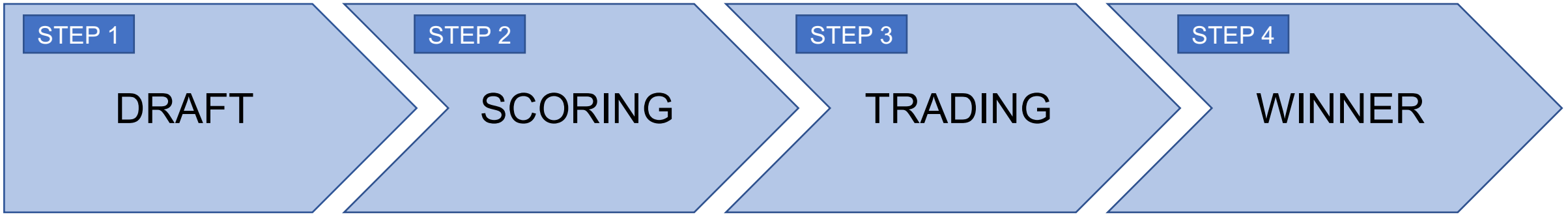
Project Summary

- What is Fantasy Football?
- Objective and Assumptions
- Data Structure and Feature Selection
- Results
- Where to go?

What is NFL Fantasy Football?

- Fantasy Football is a game where the participants serve as the general managers of virtual professional football teams. ($7 Billion Market)
- The participants choose their team roasters by participating in a draft in which all players of a real football league are available.

- Procedure



- QB, WR, RB, TE, …
- 6 for a touchdown
- 1 for 10 rushing yards
- Injured players
- 11 W 6 L **VS** 9 W 7 L

My objective is to predict fantasy points based on the historical data

- Week 17 of 2017 – Fantasy Points of Quarterbacks

Rank	Player	Opp	Passing			Rushing		Receiving		Misc		Fum	Fantasy
			Yds	TD	Int	Yds	TD	Yds	TD	FumTD	2PT	Lost	Points
1	Philip Rivers QB - LAC	OAK	387	3	-	5	-	-	-	-	-	-	27.98
2	Matthew Stafford QB - DET	GB	323	3	-	2	-	-	-	-	1	-	27.12
3	DeShone Kizer QB - CLE	@PIT	314	2	1	61	-	-	-	-	-	-	24.66
4	Jameis Winston QB - TB	NO	363	1	3	32	1	-	-	-	1	-	23.72
5	Andy Dalton QB - CIN	@BAL	222	3	-	-2	-	-	-	-	-	-	20.68
6	Russell Wilson QB - SEA	ARI	221	2	-	36	-	-	-	-	-	-	20.44

My response variable

Assumptions

- Each player's performance is dependent on the players' talent and the team's capacity.
- Each player's performance is related to the player's fundamental talent and temporary condition.
- Each team's performance is related to the team's fundamental capacity and temporary condition.

Performance

- player's talent
  - fundamental: last year
  - temporary: last 3 weeks
- team's capacity
  - fundamental: last year
  - temporary: last 3 weeks

Data are divided into three categories based on the features

- NFL Data Source: https://www.footballdb.com/fantasy-football/index.html

1. QB, WR, RB, TE

- ✓ Fantasy Points
- ✓ Passing Attempts
- ✓ Rushing Attempts
- ✓ Receiving Yards
- ✓ Fumbles Lost
- ✓ Touchdown Receptions
- ✓ …

2. K

- ✓ Fantasy Points
- ✓ PAT Attempts
- ✓ PAT Made
- ✓ FG Attempts
- ✓ 50 Yards FG Made
- ✓ …

3. DST

- ✓ Fantasy Points
- ✓ Touchdowns
- ✓ Interceptions
- ✓ Safeties
- ✓ Fumble Recoveries
- ✓ Passing Yards Allowed
- ✓ …

Additional features are constructed by averaging the last three weeks

- Pass Attempts of Quarterbacks in the 4th week of 2011

pos	year	week	player	Pass Attempts_1wk	Pass Attempts_2wk	Pass Attempts_3wk
QB	2011	4	Aaron Rodgers	38	30	35
QB	2011	4	Michael Vick	23	28	32
QB	2011	4	Cam Newton	34	46	37
QB	2011	4	Tarvaris Jackson	31	NaN	NaN
QB	2011	4	Tony Romo	36	33	36

pos	year	week	player	Pass Attempts_1wk	Pass Attempts_2wk	Pass Attempts_3wk	Pass Attempts_avg
QB	2011	4	Aaron Rodgers	38	30	35	34.33
QB	2011	4	Michael Vick	23	28	32	27.67
QB	2011	4	Cam Newton	34	46	37	39.00
QB	2011	4	Tarvaris Jackson	31	NaN	NaN	NaN
QB	2011	4	Tony Romo	36	33	36	35.00

Data Structure

- Final data structure before analysis

pos	# of Features	# of Observations
QB	84	1988
WR	84	4517
RB	84	3781
TE	84	2313
K	51	2079
DST	31	2668

pos	year	week	player	team	Fantasy Points	Fantasy Points_prev	Pass Attempts_prev
QB	2011	4	Aaron Rodgers	GB	53.0	27.0	38.0
QB	2011	4	Michael Vick	PHI	33.0	8.0	23.0
QB	2011	4	Tarvaris Jackson	SEA	27.0	12.0	31.0
QB	2011	4	Tony Romo	DAL	25.0	8.0	36.0
QB	2011	4	Josh Freeman	TB	25.0	12.0	32.0

Select features based on the p-value of 0.02

- Select useful features based on statistical significance

pos	# of Features	# of Observations
QB	84	1988
WR	84	4517
RB	84	3781
TE	84	2313
K	51	2079
DST	31	2668

pos	# of Features	# of Observations
QB	32	1988
WR	23	4517
RB	31	3781
TE	15	2313
K	13	2079
DST	12	2668

QB: Fantasy Points, Pass Attempts, Pass Completions (1 week)
Fantasy Points, Pass Attempts, Pass Completions (3 weeks)
Rushing Yards Allowed, Passing Yards Allowed (1 week)

K: PAT Attempts, PAT made (1week),
Fantasy Points, PAT Made, FG Attempts, FG Made (1 year)

Divide the data into a train set and a test set

- Train Set: 2011 – 2015 **&** Test Set: 2016 – 2017

pos	year	week	player	team	Fantasy Points	
QB	2011	4	Aaron Rodgers	GB	53	Train Set
QB	2011	4	Michael Vick	PHI	33	
QB	2011	4	Tarvaris Jackson	SEA	27	
			~			
QB	2015	17	Drew Stanton	ARI	-1	
QB	2015	17	Tom Brady	NE	5	
QB	2015	17	Blake Bortles	JAX	4	
QB	2015	17	Teddy Bridgewater	MIN	1	
QB	2016	4	Ben Roethlisberger	PIT	44	Test Set
QB	2016	4	Matt Ryan	ATL	43	
QB	2016	4	Derek Carr	OAK	31	

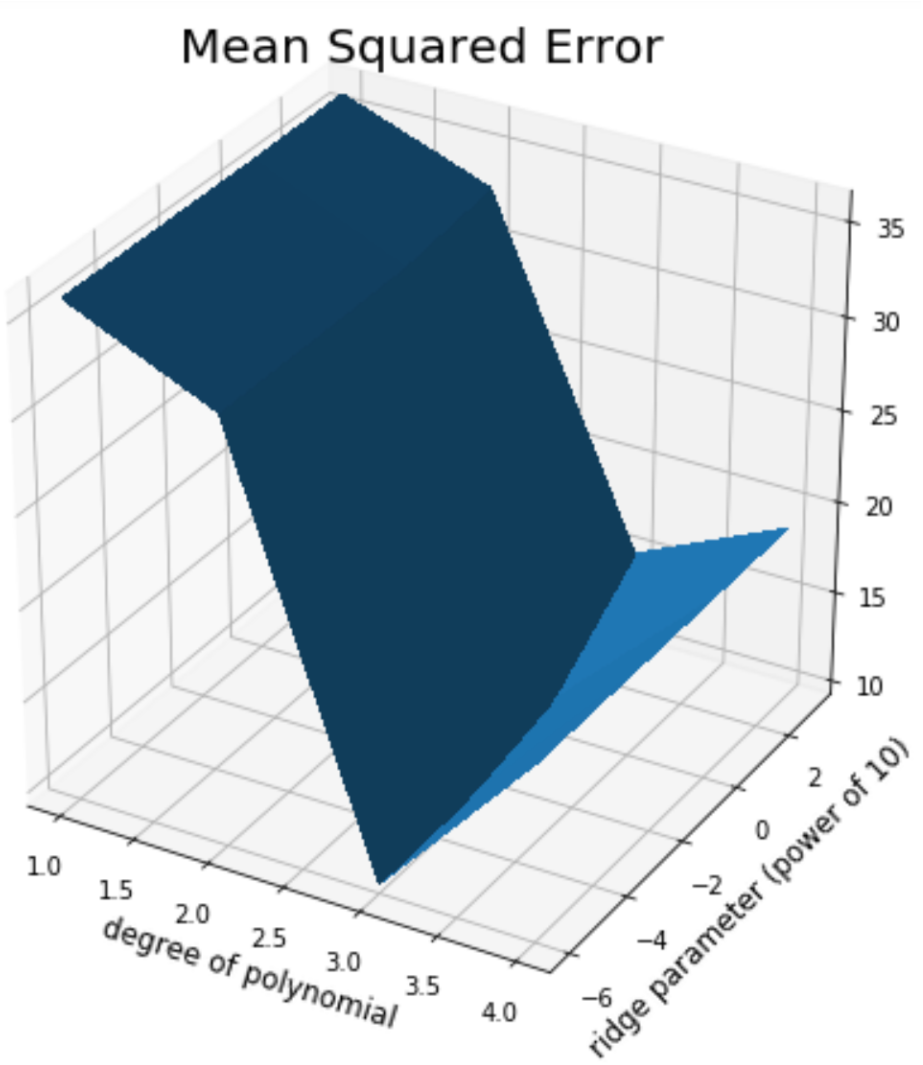
Find optimal degrees and parameters (alphas) for regularization

- Ridge Regularization

$$J(\beta) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta^T x_i)^2 + \alpha\left|\left|\beta\right|\right|_2$$

- Lasso Regularization

$$J(\beta) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta^T x_i)^2 + \alpha|\beta|$$



WR Opt Model:

Degree: 3
Alpha: 10^{-6}

Ridge is better in terms of MSE, while Lasso is better at collinearity

- Mean Squared Error?

  Ridge MSE < Lasso MSE

- Multicollinearity!

  Ridge Multicollinearity > **Lasso Multicollinearity**

Ridge is better in terms of MSE, while Lasso is better at collinearity

- Mean Squared Error?

  Ridge MSE < Lasso MSE

- Multicollinearity!

  Ridge Multicollinearity > **Lasso Multicollinearity**

- **Lasso**: number of features > 20 **&** **Ridge**: number of features <= 20

pos	# of Features	# of Observations	Model
QB	32	1988	Lasso
WR	23	4517	Ridge
RB	31	3781	Lasso
TE	15	2313	Ridge
K	13	2079	Ridge
DST	12	2668	Ridge

Results

- MSE and R-squared by positions

Pos	# of Features	# of Obs	(Model, deg, alpha)			MSE	R2
QB	32	1988	Lasso	4	10^{-6}	0.2	10%
WR	23	4517	Ridge	3	10^{-6}	9.8	8%
RB	31	3781	Lasso	4	10^{-6}	4.9	9%
TE	15	2313	Ridge	3	10^{-6}	10	6%
K	132	2079	Ridge	3	10^{-3}	10.7	3%
DST	12	2668	Ridge	4	10^{-6}	2.2	10%

Better than 4%, but …

Results: Bow-Cox transformation

- Box-Cox transformation

$$y \rightarrow \frac{y^{\lambda} - 1}{\lambda} \quad (y: fantasy\ point)$$

Pos	# of Features	# of Obs	(Model, deg, alpha)			MSE	R2	MSE_pow	R2_pow
QB	32	1988	Lasso	4	10^{-6}	0.2	10%	0.1	11%
WR	23	4517	Ridge	3	10^{-6}	9.8	8%	2.3	8%
RB	31	3781	Lasso	4	10^{-6}	4.9	9%	1.6	9%
TE	15	2313	Ridge	3	10^{-6}	10	6%	2.1	6%
K	132	2079	Ridge	3	10^{-3}	10.7	3%	4.7	3%
DST	12	2668	Ridge	4	10^{-6}	2.2	10%	0.6	11%

Where to go?

- Other features: Home Stadium? Opponents (Who is marking my QB?)
- Different Models: Random Forest?
- Filtering: Outliers

THANK YOU

Quenstions?