Predict Player Performance for an Optimal Fantasy Game Lineup
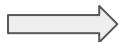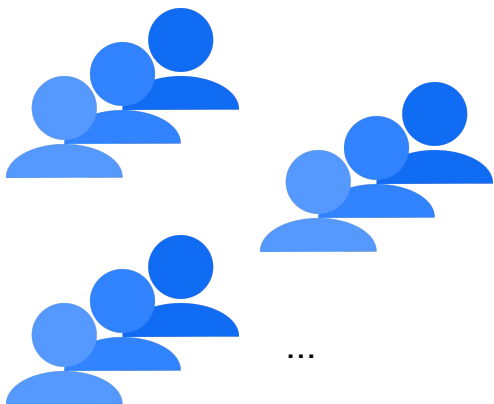
Tian Sang, Zhaoxi Wang
11/04/2019

Introduction

DRAFT KINGS
THE GAME INSIDE THE GAME.

Create an 8-player lineup!

X8

Constriants like Salary cap, Team, Position

Before match day

player 1:   52
player 2:   14
player 3:   33
...

Performance score calculated by a given equation.

lineup 1: X8   300
lineup 2: X8   296
lineup 3: X8   307
...   ...

After match day

Methods

Predict player score.

- Collect historical player statistics
- Data transformation & feature engineering
- Modeling

Form lineup given constriants.

- Generate Lineup by linear programming
- Diversity Constriants:

Predict player score.

Collect historical player statistics: stats.nba.com

#	PLAYER	GP	MIN	PTS	FGM	FGA	FG%
1	James Harden	78	36.8	36.1	10.8	24.5	44.2
2	Paul George	77	36.9	28.0	9.2	21.0	43.8
3	Giannis Antetokounmpo	72	32.8	27.7	10.0	17.3	57.8
4	Joel Embiid	64	33.7	27.5	9.1	18.7	48.4
5	Stephen Curry	69	33.8	27.3	9.2	19.4	47.2

Data transformation & feature engineering

- Time series features, e.g. "previous game points", "2nd previous game points" etc.

- Encode categorical feature team: winning rate, average points

Predict player score.

H2OAutoML(max_models = 30, max_runtime_secs=150, seed = 1)

Modeling: H2O autoML

- Automating training & tuning of hyper-parameters
- Often has good performance in practical.
- Returns a leaderboard with the best model.

Model 1
Model 2
…
Model 30

Best model

Stacking of all models has the lowest MAE!

model_id	mean_residual_deviance	rmse	mse	mae
StackedEnsemble_AllModels_AutoML_20191029_185211	97.465	9.87244	97.465	7.79825
GLM_grid_1_AutoML_20191029_185211_model_1	97.76	9.88736	97.76	7.80035
StackedEnsemble_BestOfFamily_AutoML_20191029_185211	97.8783	9.89335	97.8783	7.81343
XGBoost_grid_1_AutoML_20191029_185211_model_2	99.8868	9.99434	99.8868	7.88746
GBM_5_AutoML_20191029_185211	100.385	10.0192	100.385	7.9257

Form lineup given
constriants.

Generate Lineup by linear programming

FEASIBLE REGION
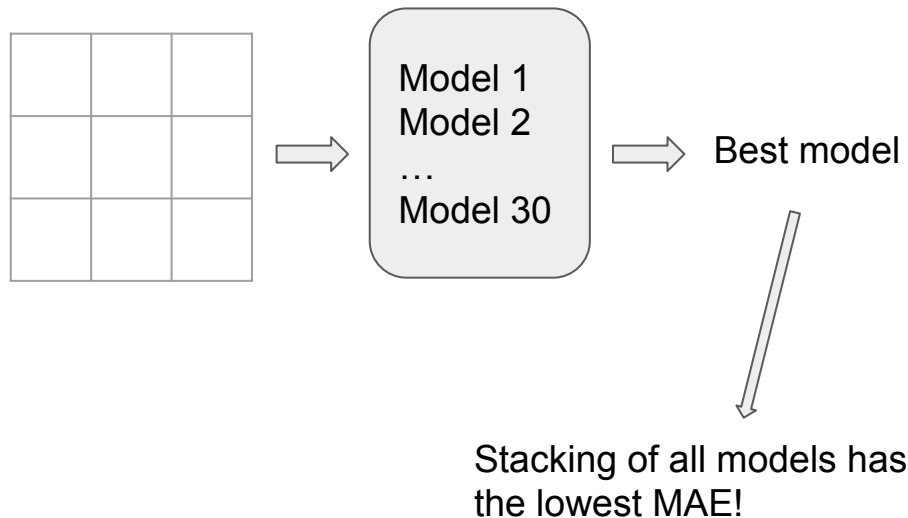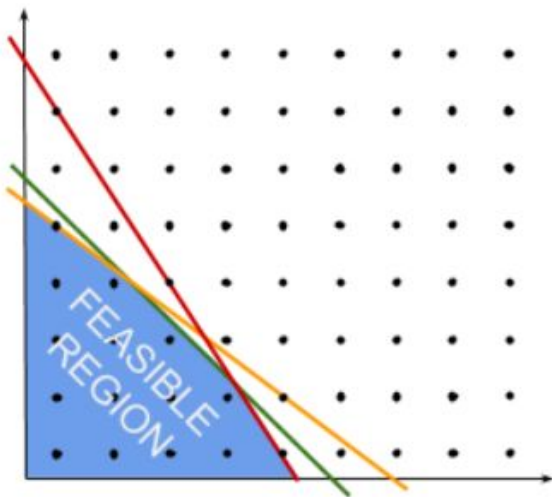
Goal:
$$\sum_{p=1}^{N} f_p x_{pl}$$

Subject To:
$\sum_{j=1}^{p} c_j x_{ij} \leq 5000,$ (budget constraint)
$\sum_{k=1}^{p} x_{ij} = 8,$ (lineup size constraint)

position constriants:

$1 \leq \sum_{j \in C} x_{ij} \leq 2$
$2 \leq \sum_{j \in (PG,SG)} x_{ij} \leq 3$
$2 \leq \sum_{j \in (PF,SF)} x_{ij} \leq 3$

team constriants:

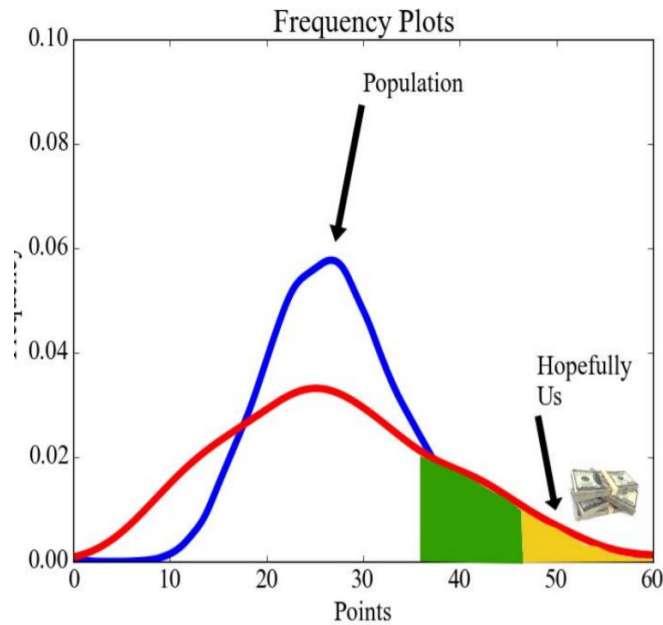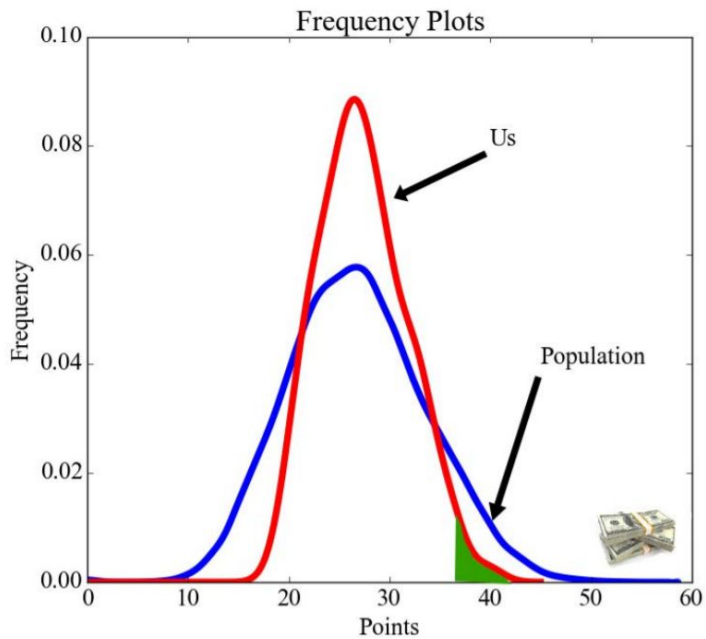$t_{il} \leq \sum_{k \in T_l} x_{ik}, \quad l = 1, \ldots, N_T$
$\sum_{l=1}^{N_T} t_{il} \geq 3$
$t_{il} \in \{0,1\}, \quad l = 1, \ldots, N_T$

Form lineup given constriants.

Diversity Constriants:
Make sure lineup i has only a few overlap players from lineup 1 to i-1

$$\sum_{p=1}^{N} x_{pk}^* x_{pl} \le \gamma, k = 1, \ldots, l-1$$

Frequency Plots

Us

Population

Frequency Plots

Population

Hopefully Us

Evaluation

Sell lineup strategy ($30 monthly subs fee 😔 expensive!)

draftkings_NBA_2019-02-01_players.csv
draftkings_NBA_2019-02-02_players.csv
draftkings_NBA_2019-02-03_players.csv
draftkings_NBA_2019-02-04_players.csv
draftkings_NBA_2019-02-05_players.csv
draftkings_NBA_2019-02-06_players.csv
draftkings_NBA_2019-02-07_players.csv
draftkings_NBA_2019-02-08_players.csv
draftkings_NBA_2019-02-09_players.csv
draftkings_NBA_2019-02-10_players.csv
draftkings_NBA_2019-02-11_players.csv

VS. Our prediction ⟹ Avg. MAE:

FC's	Ours
7.80	7.67

Feb. 1 - Mar. 31, 2019

Evaluation

- Simulation on Feb.01
- 4 of our 49 lineups won

Conclusion

- A capable machine learning solution for selecting lineups
- Acceptable performance as a prototype
- Need to improve ROI in the future