Draft Kings and Queens: Predicting the Optimal Fantasy Basketball Team

Eric Hermann and Adebia Ntoso

CS 229: Machine Learning

Stanford University, Department of Computer Science

Parameter Informativeness

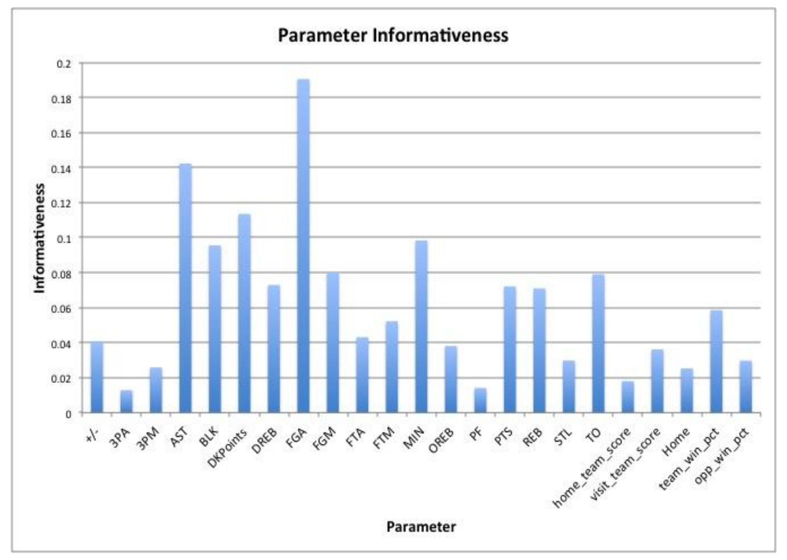
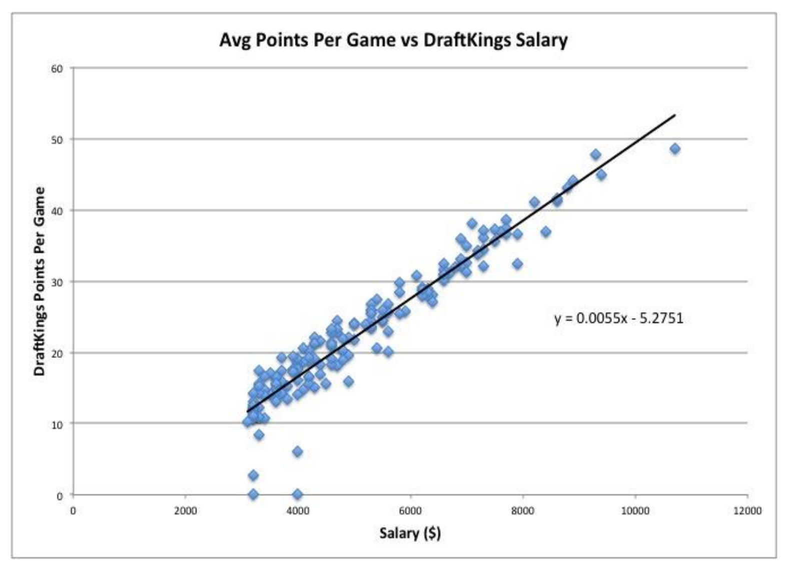


Figure 1: Relative informativeness of particular parameters of the regression algorithm. Interestingly, categories like field goals attempted and assists are the most informative, meaning they are more directly correlated to future DraftKings points.

Figure 2: Each player's average points per game vs their DraftKings salary cost. The relationship between DraftKings salary and their points per game is relatively linear, which gave justification for using linear regression as a tool to solve the problem



Regression

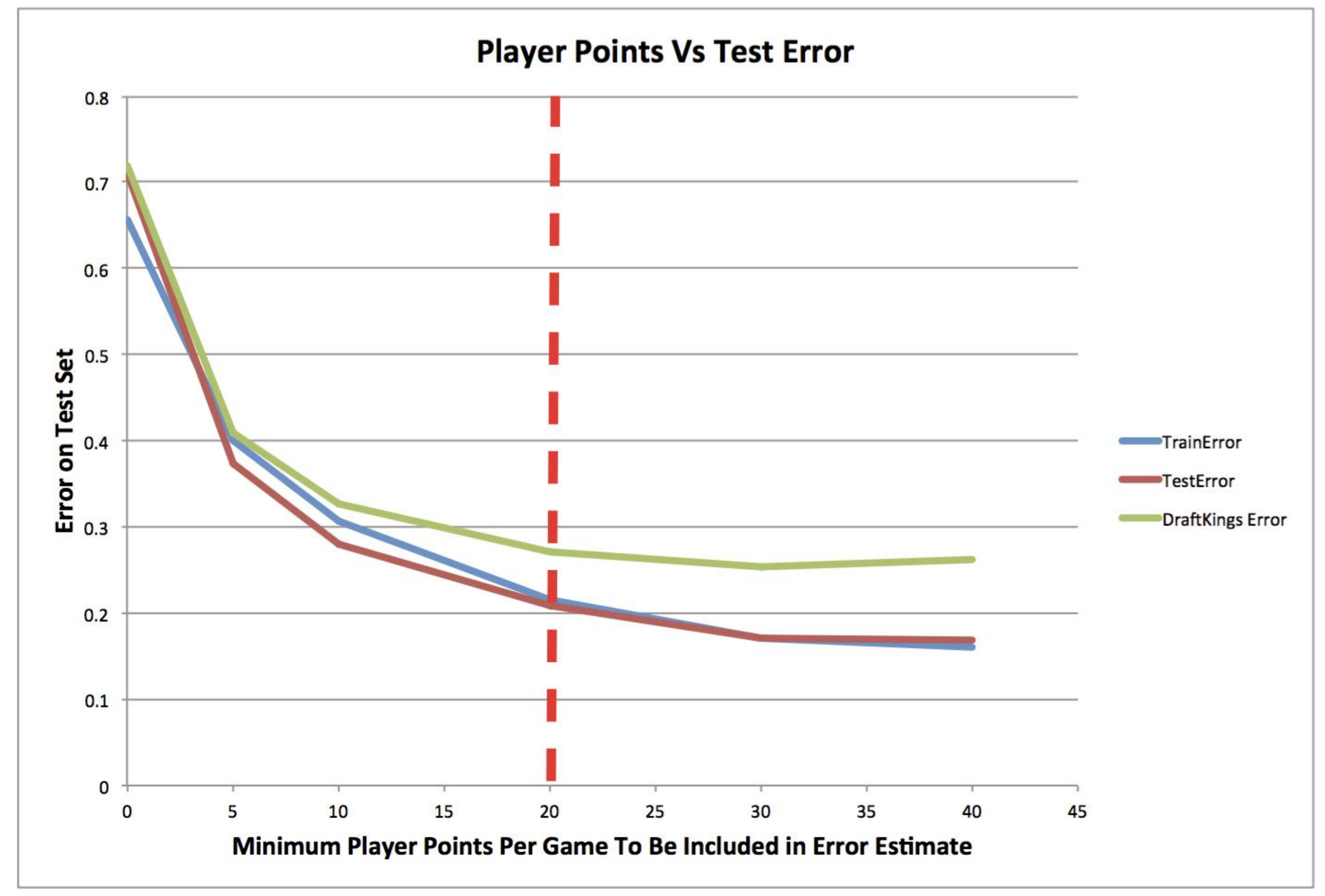


Figure 3: Train and test error of the regression algorithm as compared to the error of DraftKings in prediction. Most players that end up being drafted have scores higher than 20 points, at which point the regression algorithm is significantly more informative than the DraftKings algorithm.

Future Regression Work

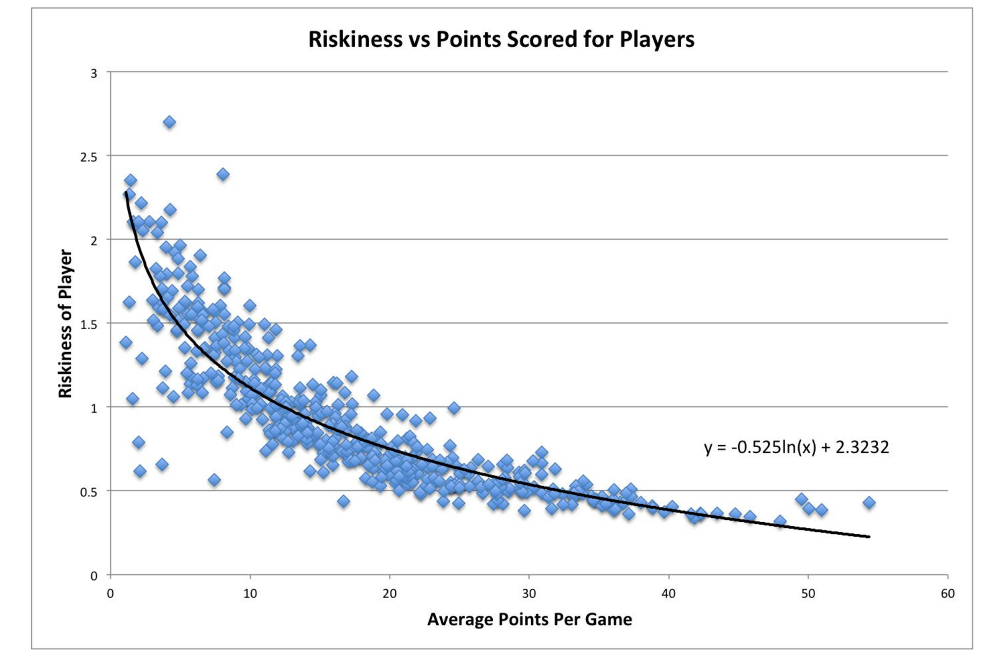


Figure 5: Player variance/mean divided by average variance/mean gives a measure of "riskiness" of a player. More risky players are more likely to win fantasy sports if the number of teams created is high, because they have a higher upside. A more complex measure of riskiness is necessary to factor this into the regression problem.

Figure 4: Discretization of the data has little effect on test error.

Number of Discretizations	1	2	3	4
Test Error	0.279838	0.278183	0.27795	0.277933
Discretization Points	10	10,30	10, 20, 30	10,20,30,40

Constraint Satisfaction Problem

...

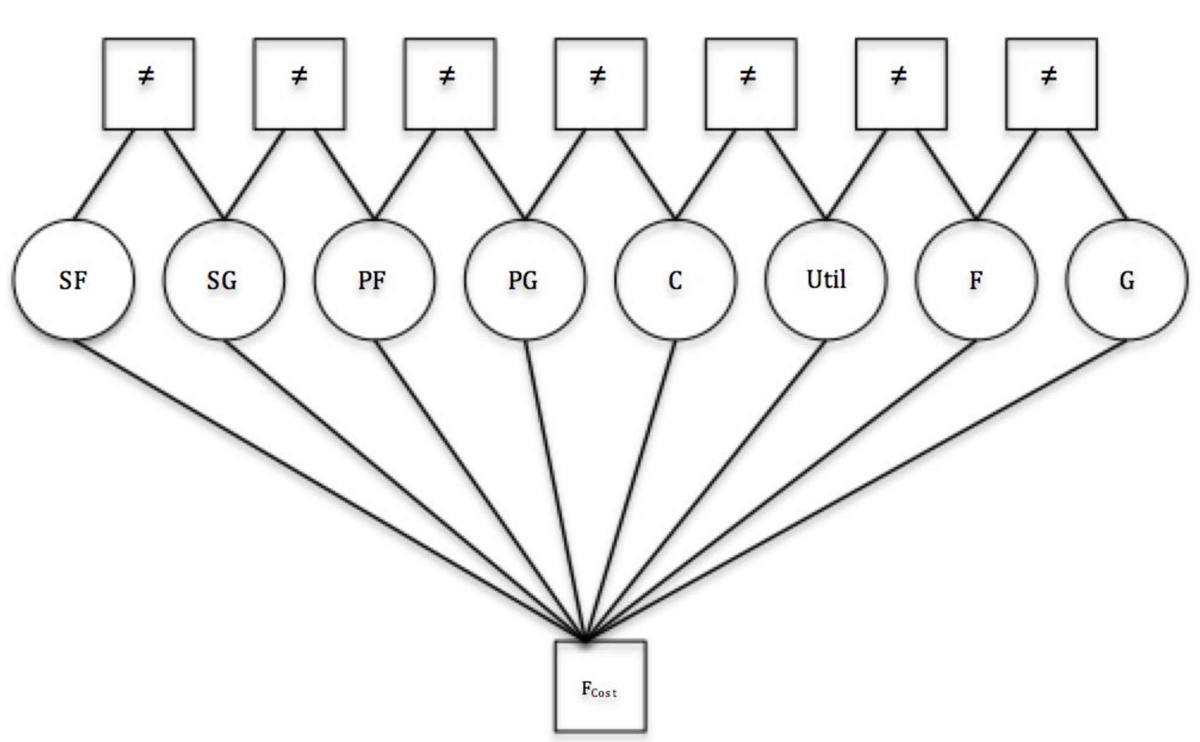


Figure 6: This is a detailed model of our CSP. The variables for our search problem are the positions to which a player must be assigned. There are two sets of factors in our graph. The first is a set of factors indicating that a player cannot be assigned to two positions. For every i, j:

$$f(X_i=x_i, X_j=x_j) = 1\{ x_i.name() \text{ != } x_j.name() \}$$

The second is a set of factors to restrict the total cost of the team to \$50,000. This is done by the use of auxiliary variables since each factor can only depend on two variables. The effective resulting factor is:

$$f(X_1=x_1, \ldots, X_n=x_n) = 1\{ x_1.cost() + \ldots + x_n.cost() \leq \$50,000 \}$$

Backtracking Search

Optimal Team	PG: George Hill PF: Kenneth Faried SG: Victor Oladipo SF: Tobias Harris C: Marc Gasol Util: Rudy Gobert G: Will Barton F: Zach Randolp
Score	242.239
Cost	$50,000
Runtime	80.8438541889

Figure 7: It is computationally impossible to run backtracking search on the CSP for the dataset of all the players. This this algorithm increases exponentially with respect to n where n is the number of players to consider. As a result, our algorithm would take $O(n^8)$ time to complete. We ran backtracking search on an player dataset of only 35 players and found the optimal assignment displayed above.

Beam Search

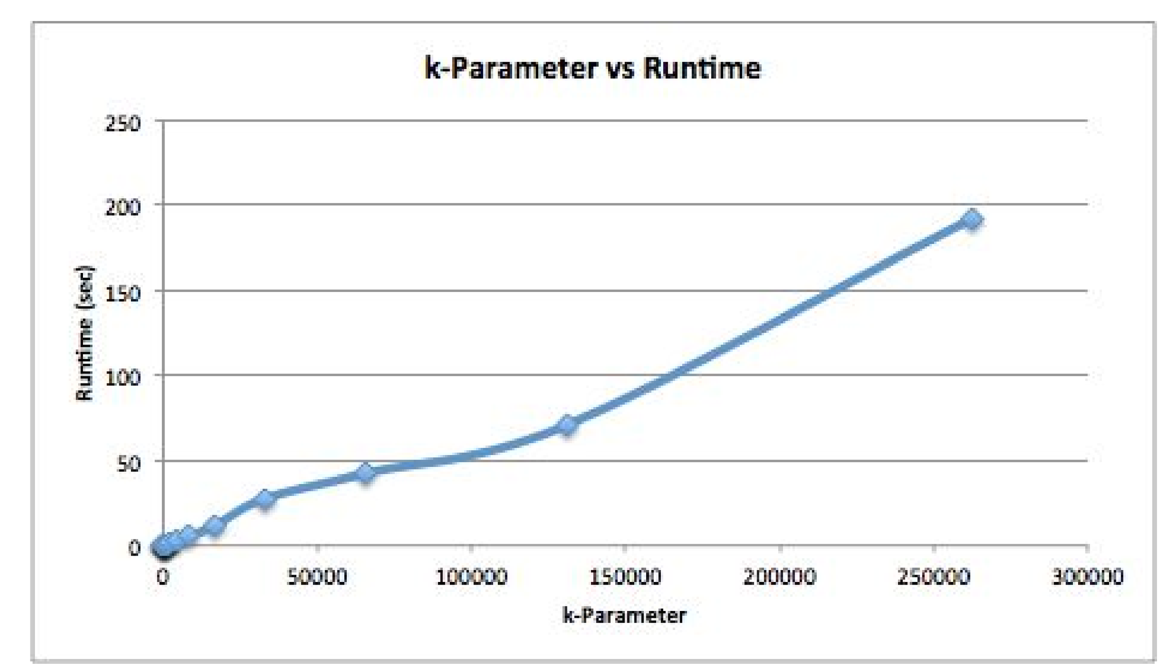


Figure 9: Varying the k-parameter of our beam search problem in turn varied the runtime of the algorithm. Although beam search isn't required to find to optimal assignment, it found a reasonable assignment with a high predicted score relatively quickly. The runtime to increased linearly with respect to k which matched the expected runtime of O(n(KB)log(KB))

Figure 8: The top 3 predicted teams produced our beam search solver.

Team	Richard Jefferson, Marc Gasol, Draymond Green, Randy Foye, Rodney Stuckey, Gordon Hayward, Victor Oladipo, LeBron James	Victor Oladipo, Marc Gasol, Draymond Green, Randy Foye, Ben McLemore, Joe Johnson, Alec Burks, LeBron James	Trevor Booker, Marc Gasol, Draymond Green, Randy Foye, Rodney Stuckey, Omri Casspi, Victor Oladipo, LeBron James
Score	280.486	272.153	271.040
Cost	$50,000	$49,400	$49,400