Avalanche: Algorithm Optimization

Mario Aguilar, Margaret Baxter, Patrick Quach,
Christine Russell

Optimizing Power Rankings

Based solely on a team's Average Score Margin- try initializing ASM with all data from previous seasons? What is the difference when ASM's are and are not initialized based on the preseason? (The early season could be more heavily weighted to predict for the preseason data (still updated as the season progresses) and later games could put more emphasis on the regular season only ASMs). This will need to be a generic optimization- not just 2015, as it will be used for optimizing Bayes theorem.

Original Power Rank Results

Year	Accuracy
2015	44.0625
2014	61.4611673
2013	66.3340983
2012	64.3685642
2011	60.6757354

First Optimization

Year	Accuracy
2015	55.46875
2014	57.421875
2013	56.25
2012	62.109375
2011	53.515625

First Optimization: In fact a bug fix, all teams must be referenced against their weekly opponents.
Results: Power rankings were more accurate for the 2011-2015 season when only ranking teams against the teams they played in the pre-season. This suggests either some underlying logical connection that is not inherently clear, or, more likely, power rankings are not very accurate in their base form.

Optimizing Bayes

Use Bayes' to compute the probability of win given the team's power rankings.

- Power Ranks must be computed for every team in the testData database and change over the course of the seasons- use best predictor from Optimized PRs and one base case regular PR.

Original Bayes Results

Year	Accuracy
2015	52.5195312
2014	63.2691147
2013	61.9794057
2012	63.8651763
2011	60.1623535

- Does not account for home team, does not use all games as reference.
  This optimization uses only the previous games played by the home team, whether they were the home team or the away team, and computes the likelihood that this team will win. Also takes into account their opponent, but does not perform any additional statistical analysis based solely on their opponent; everything is related back to the home team.

My Bayes Results: Optimization 1

Year	Accuracy
2015	70.3125
2014	69.53125
2013	78.515625
2012	73.046875
2011	68.75

- Does not account for home team, does not use all games as reference, includes a built-in skew that favors home team.
  This optimization uses only the previous games played by the home team, whether they were the home team or the away team, and computes the likelihood that this team will win. Also takes into account their opponent, but does not perform any additional statistical analysis based solely on their opponent; everything is related back to the home team. The statistical analysis is also slighlty in favor of the home team winning.
  Results: Only slightly more accurate than optimization 1, and not reliably so.

My Bayes Results: Optimization 2

Year	Accuracy
2015	70.703125
2014	71.875
2013	76.953125
2012	73.4375
2011	66.015625

- This Bayes selection references the cases the given home team was previously also playing home team. This does not take into account cases where this team was also the away team as that would give their opponent the advantage. This only ranks the likelihood of the home team to win, no data is computed about the away team although the specific case of previous games where both teams played each other is taken into account.
  Results: Much better accuracy than with the built in skew.

My Bayes Results: Optimization 3

Year	Accuracy
2015	73.82815
2014	75.78125
2013	78.125
2012	80.46875
2011	75

- This Bayes selection is even more particular to the home team, only cases where the given home team was previously also playing home team are used at all. This does not take into account cases where this team was also the away team as that would give their opponent the advantage. This only ranks the likelihood of the home team to win, no data is computed about the away team although the specific case of previous games where both teams played each other is taken into account.
  Results: Best accuracy so far, allowing Bayes to select in such a manner that maintains the integrity of a home team advantage has proved to be the most efficient means of calculating the home team's chance of victory.

My Bayes Results: Optimization 4

Year	Accuracy
2015	82.8125
2014	83.984375
2013	82.8125
2012	83.59375
2011	78.125

Optimizing Pi Rating

The original Pi-Rating algorithm uses a predicted goal difference to determine a winner. Each team has a home and away rating, which relate to their predicted home or away score differential. These ratings correlate to the performance that can be expected when playing home or away. Some teams only perform well when they play at home, and so forth. Originally, the Pi-Rating algorithm sought to predict the actual point difference between two teams. When selecting a team, the predicted home score value for the home team was compared with the predicted away score value for the away team. Once a winner is predicted, the values must be updated for the next week's simulation. The actual game differential is entered into an error function, $\Phi(error)$, to determine how to update the teams' statistics. Three other constants are used:

$$\lambda = 0.0247678216662108$$
$$\gamma = 2.50660814461606 \times 10^{-5}$$
$$C = 3.3936218202084403$$

However, the accuracy given when comparing the teams' predicted score differential was very close to 50%.

Original Pi-Rating

Year	Accuracy
2015	51.953125
2014	55.859375
2013	59.765625
2012	56.250000
2011	57.421875

In order to optimize the Pi-Rating prediction, the teams' home and away ratings were used instead. This gave the following improvement in accuracy.

Optimized Pi-Rating

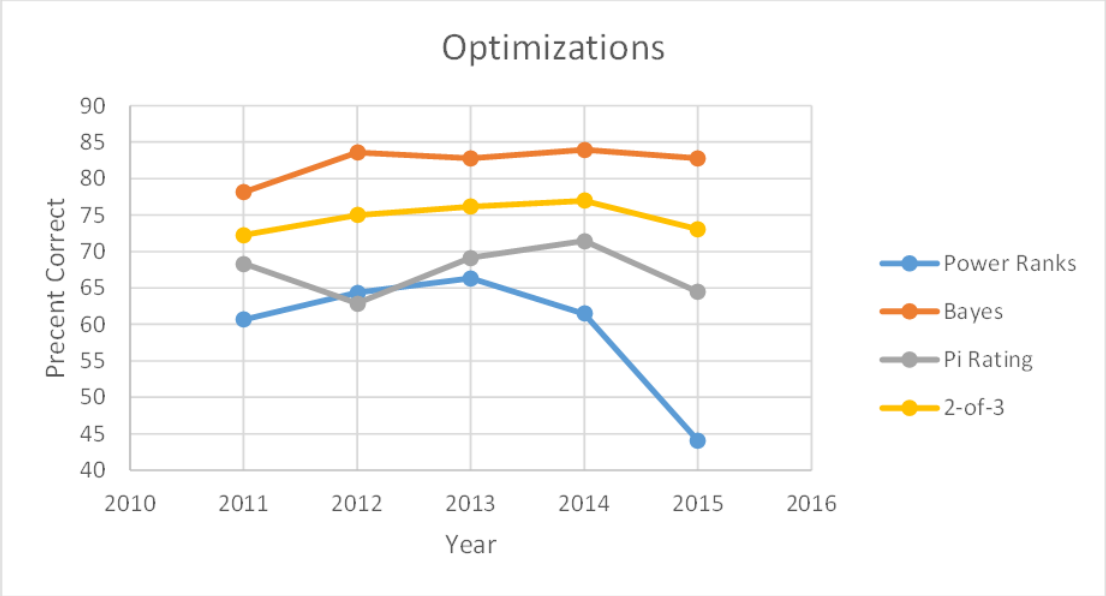
Year	Accuracy
2015	64.453125
2014	71.484375
2013	69.140625
2012	62.890625
2011	68.359375

Overall Optimizations

In an attempt to use all three prediction services without compromising accuracy, we developed a 2 out of 3 basis for selecting predictions. Simply put, the prediction that was shared by two or all of the predictors is the one selected by this algorithm.

2-of-3 Selector

Year	Accuracy
2015	73.046875
2014	76.9531525
2013	76.171875
2012	75.0
2011	72.265625



The Big Winner predictor utilizes the 2-of-3 selector to predict the outcomes of NFL games in our android application. Even though this is not the most accurate, the optimized bayes 4 algorithm places too much emphasis on the actual recorded game, in other words, knowing the actual outcomes of the games skews bayes' results.