Matthew Grant

ME 395- Mechanistic Data Science

October 5, 2021

Mechanistic Data Science Proposal

One example of a real life problem which can be solved with MDS is determining what kinds of skills and statistics are necessary to win the MVP in the NBA. This would be a type I problem since there is plenty of available data collected on the performance of NBA players, although there is no mathematical theory correlating this data to likelihood to win Most Valuable Player. I selected basketball as the sport of interest for a couple reasons. Firstly, there are only 5 positions on a basketball team. This makes it easier to discriminate by position since the player’s role on the team most likely has a large impact on their likelihood of winning MVP. Another reason why I chose basketball is because each team plays all the other teams at least twice per season. This helps eliminate sampling bias from certain players avoiding “tougher” teams through scheduling. A third reason is that basketball is a very popular sport and therefore statistics are accurately tracked and verified.

The six steps of MDS:

1. Data collection and generation: There are several NBA databases that are available. Basketball Reference is the most popular of these databases and contains statistics from every NBA game ever played since 1947.
2. Mechanistic feature extraction: The NBA tracks several statistics each game including points scored, rebounds, assists, steals, blocks, turnovers, fouls, and shots attempted. From these base stats, more advanced statistics can be calculated such as shooting percentage and offensive win shares. All of these statistics are available for every player on Basketball Reference and they can be averaged over any particular season or career.
3. Dimension reduction: Since not every measurable stat is critical to winning MVP, we can narrow it down ones we determine are critical
4. Reduced order modeling: Use these reduced dimensions to compare leaders in certain statistics to the amount of MVP votes received in any particular year
5. Mechanistic learning through regression: Knowing who the previous MVP winners are, and their statistics in their winning season, we can build a model to see the correlation between our key stats and the winners. (Did the MVP lead the league in points scored? If not, did the points scored leader receive any MVP votes?)
6. System and design: Complete NBA MVP tracker. With the input of yearlong averages of relevant statistical categories of several players, the model will be able to rank those players in terms of likelihood to win MVP.