INFSCI 0530 Decision Making in Sports

Konstantinos Pelechrinis
kpele@pitt.edu

Course Description Data and analytics have been part of the sports industry from as early as the 1870s, when the first boxscore in baseball was recorded. However, it is only recently that advanced data mining and machine learning techniques have been utilized for facilitating the operations of sports franchises. While part of the reason is related with the ability to collect more fine-grained data, an equally important factor for this turn to analytics is the huge success and competitive advantage that early adopters of investment in analytics enjoyed (popularized by the best-seller ``Moneyball'' that described the success that Oakland Athletics had with analytics). Draft selection, game-day decision making and player evaluation are just a few of the applications where sports analytics play a crucial role today. Apart from the sports clubs, other stakeholders in the industry (e.g., the leagues' offices, media, etc.) invest in analytics. The leagues increasingly rely on data to decide on potential rule changes. In this course, we will introduce data science concepts for sports analytics. Students will get introduced concepts related to data collection, data quality, data analysis and modeling as well as data visualization.

Learning outcomes

Upon successful completion of the course students will be able to:

- Understand how to interpret and use probabilities both in the field of sports decision making, as well as, in general.
- Understand the Bayesian process.
- Become familiar with basic data analysis techniques, such as linear and logistic regression.
- Obtain a gentle introduction to Monte Carlo simulations.
- Understand the notion of overfitting and how you can avoid it.
- Understand what comprises a useful visualization and why effective communication is crucial.

Reference Textbook (optional): W. Winston, “Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball and Football”, Princeton University Press, 2009.

Course Requirements & Grading:

The evaluation will be based on the following components:

1) Homework: 20%
2) Project: 30%
3) Midterm: 25%
4) Final: 25%

Tentative Schedule

Week	Topics covered
Week 1	Where are analytics used in sports? What is the current state-of-the art? **Data collection, quality and ethical concerns**
Week 2	**Empirical probability and statistical tests** **concepts**: two-point conversion in NFL, fourth-down decision in the NFL, end-game basketball strategy, do officials fix NBA games?, is the hot-hand real?
Week 3	**Bayesian player evaluation** **Sports Science and Analytics**
Week 4	**Introduction to regression**: Bradley-Terry model, Oliver's four factors, in-game win probability **Evaluating probability models**: Brier score, probability validation curves
Week 5	**Team rating systems**: Elo rating, regression-based rating, network-based rating **Player rating systems:** QBR, adjusted plus/minus
Week 6	**What is overfitting? How to avoid it:** regularization and its Bayesian interpretation **Occam's razor and model selection**
Week 7	Midterm
Week 8	**Introduction to Monte Carlo simulations:** simulating a tournament **What is resampling?**
Week 9	**Sports and behavioral economics:** decision making and risk aversity **Natural experiments:** the case of corner three shots
Week 10	**Why are visualizations important?** **Principles of visualizations**
Week 11	**Introduction to Game Theory:** (mixed) strategies, Nash Equilibrium **Case studies:** passing vs rushing in NFL, penalty kicks in soccer, defending the corner three in the NBA
Week 12	**Unsupervised learning** **Clustering:** grouping players and teams
Week 13	Project presentations
Week 14	Final

The project can be in groups of up to 3 students. All the group members will get the same grade for the project – essentially there will be a grade for the project and not the team members. So, you will have to make your first decision on if and who you want as your teammates (i.e., build your roster). The project can be of three separate types:

1. A "traditional" analytics project, where the goal is to use data to build a predictive or descriptive model to answer a particular question. This will involve data exploration, data cleaning, data modeling, model selection and visualizations.

2. A survey of a specific analytical method and how it can/has been used in (decision making in) sports
3. A combination of the two. E.g., how have people tackled a decision that needs to be made frequently in sports? For instance, what are the approaches that people have used for making decisions on 4th down in the NFL. What are some expected models that have been developed and how have they evolved over time. What are the pros and cons of each?

University Policies:

Collaboration Policy

Miderm: No collaboration

Final: No collaboration

Homework: No collaboration. However, discussion is encouraged, but stay away from one another answers. When discussing a problem, avoid taking notes. If you filter the discussion through your own memory, you are very unlikely to submit answers that "accidentally" indicates you collaborated. This is for your own protection.

Material Covered

You are responsible for all material covered in lectures, as well as, reading and homework assigned.

Academic Integrity

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

G-grade Policy

No G-grade will be granted and students must finish their work load during the semester.

Disability Services

If you have a disability that requires special testing accommodations or other classroom modifications, you need to notify both the instructor and Disability Resources and Services no later than the second week of the term. You may be asked to provide documentation of your disability to determine the appropriateness of accommodations. To notify Disability Resources and Services, call (412) 648-7890 (Voice or TTD) to

schedule an appointment. The Disability Resources and Services office is located in 140 William Pitt Union on the Oakland campus.

Copyright Notice

Course materials may be protected by copyright. United States copyright law, 17 USC section 101, et seq., in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials. See Library of Congress Copyright Office and the University Copyright Policy.

Statement on Classroom Recording

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.