INFSCI1091 Moneyball 2.0: Winning in Sports with Data

Data Competition

As part of this course there will be a data competition which is optional and it will be for extra credit. The topic of the competition is **NBA game predictions**. In particular, you will be required to submit a win probability for the home team of all the regular season NBA games between March 15th, 2018 and April 11th, 2018 (end of regular season). I will provide you with some basic data, but you are allowed to use any dataset you think it will help your predictions.

Deliverables

You are expected to provide two deliverables:

1. A csv file with 4 columns:
    a. Game Date (format: Month Day, Year – e.g., Mar 20, 2018)
    b. Home Team (format: full name – e.g., Boston Celtics)
    c. Away Team (format: same as above)
    d. Win probability (format: decimal number between 0 and 1)
2. A short report (maximum 2 pages) on the data and methods you used for the prediction.

Timeline

The csv file will need to be delivered by March 14, 2018, 11:59pm, while the short report should be delivered the last week of the classes.

Evaluation

The evaluation and ranking of the participants will be based on the Brier score of the predictions. Remember the lower the Brier score the better. Hence, the participant with the lower Brier score will be 1st. The maximum extra credit for a student will be 10%. The formula for the total extra points given to student i is:

$$EC_i = 10 \cdot \frac{N - r_i}{N - 1}$$

where r_i is the rank of student i and N is the number of total students that participated in the competition.