Goal: I want to predict who will win the game. (i.e. Predict 1 or 0)

Preconditions: I know who will be playing in the game for the home team and away team. I will also know each player’s performance prior to heading in the game.

Players: The players will be divided into pitchers and batters, at least to begin. The stats connected to each player type are different given the ability to measure their performance.

**Issue:** The number of parameters for batters and pitchers will be variable. Second, the number of players entering the model will be variable. How can you format the data?

Illustration:

Home Team: (Where xn are the number of features for batters and num\_players are the number of players on the bench)

1. Catcher [x1, x2 … xn]
2. First Base [x1, x2 … xn]
3. Second Base [x1, x2 … xn]
4. Third Base [x1, x2 … xn]
5. Shortstop [x1, x2 … xn]
6. Left Fielder [x1, x2 … xn]
7. Right Fielder [x1, x2 … xn]
8. Center Fielder [x1, x2 … xn]
9. Pitcher / Designated Hitter [x1, x2 … xn]
10. Bench Player [x1, x2 … xn]

Num\_players. Bench Player [x1, x2 … xn]

1. Starting Pitcher [y1, y2 … yn] \*Where yn are the number of features related to pitchers.
2. Middle Reliver [y1, y2 … yn] \* Where num\_pitchers are the number of pitchers sitting on the bench.
3. Setup Pitcher [y1, y2 … yn]
4. Closer [y1, y2 … yn]

Num\_pitchers [y1, y2 … yn]

The same set-up is consistent for the visiting team.

Questions:

1. How can I format the data where it is able to classify wins or losses?

2. Do the size of the matrices being inputted have to be the same size? Can there be variance in the size?