Luke Gerhart

Homework 2

Git repo: <https://github.com/lukegerhart/sportshw2>

The github repo contains all code and custom files I generated and used in the process of answering the questions.

(1a) In order to rank the leagues in homefield advantage, I used the cost function for regression based team ratings, but set the rating for both teams to 0. Since the *mi* is known, the function then becomes a function solving for *H*, the home field advantage. The idea is that if two teams are rated equally, then the home team’s margin of victory is due entirely to home field advantage. The function is essentially:

I then estimated *H* using a gradient based approach:



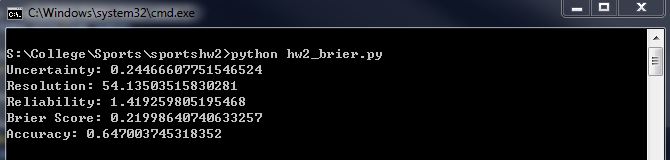
The results are:



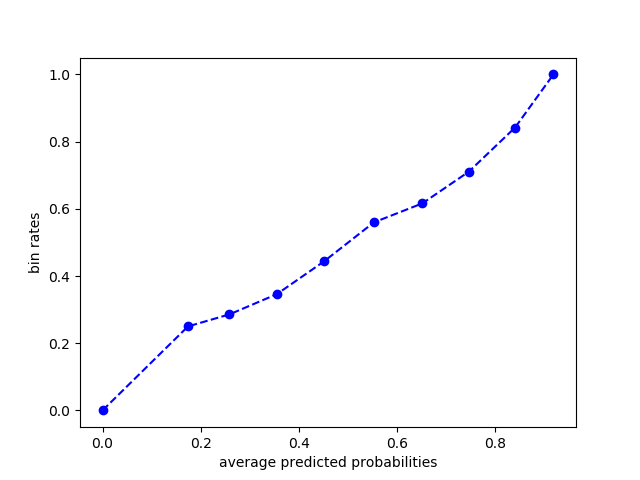
The NBA has the most advantage, then NFL, NHL, MLB.

(1b) The home field advantage for the NHL and NBA haven’t changed much. However, the NFL’s home field advantage more than tripled, going from .797 points in 2006 to 2.558 points in 2016. On the other hand, MLB’s home field advantage was halved, going from nearly 0.2 points to just .098 points.

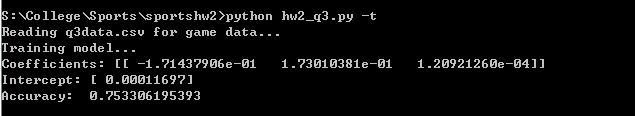
(2) See hw2\_brier.py. I forgot to prune out neutral field games from the calculations so the numbers might be a little off.



The reliability curve is below:



The bins were split by increments of .1 (0, 0.1, …, 1).

(3) I trained a relatively simple model. I chose 3 independent variables – away score, home score, and remaining time (in seconds). The model was trained on a randomly generated 75% subset of the dataset. The dataset consisted of every play from the 2 NBA seasons. 

The coefficients are the coefficients for away score, home score, time remaining respectively. The accuracy was calculated using the remaining 25% of the data set.