Computational methods in Statistics: Project proposal

Name: Yukang Xu(002462280)

Title: MVP performance prediction using Hidden Markov Model (HMM)

Description: In NBA, who is the MVP is always the interesting topic. As a basketball player, their performance is unpredictable, for example, age, co-worker, time are responsible for this. Because of its very own erratic nature, it is very difficult to forecast the future performance. We are interested in 3 states "amazing", "great" and "trash". These states determine the player value and usually invisible to the audience. They can only be estimated by the observations related to hidden states. At each time point, the HMM emits an observation and changes a state with certain probabilities. In this project, Hidden Markov Model will be used to predict the future trend of MVP. The hidden sequence of states and their corresponding probabilities will be found for a given observation sequence.

Set up: An HMM consists of following elements

S: A set of hidden states

O: A set of observations

P: State Transition Probability Matrix.

E: Observation Emission Probability Matrix.

π: Prior Probability Matrix

Then, HMM is defined as $\lambda=(P, E, \pi)$

Objectives:

1. Finding the probability of occurrence of the observation sequence $O = O_1, O_2, \ldots, O_n$ i.e. given the model λ. we want to find $P(O| \lambda)$
2. Given the observation sequence O and a model λ, how do we choose a state sequence $q_1, q_2, \ldots, q_n$, that best explains the observations. i.e. calculating the most likely sequence of hidden states that produce this observation sequence O.
3. Updating model parameters $\lambda=(P, E, \pi)$ to maximize $P(O| \lambda)$

Procedure:
We will consider points of Stephen Curry, James Harden and Giannis Antetokounmpo in their this season for three months from http://www.espn.com. We will have two values for observations, "I" and "D" for price increased and price decreased respectively. If todays close value is greater than yesterdays close values, then we will observe "I", otherwise we will observe "D".

We will use Baum-Welch algorithm to train the HMM i.e. given the observations and general structure of the HMM, we will determine the HMM parameters λ that best fit the observations. The forward-backward algorithm will be used to compute the $P(O| \lambda)$. Objective 2 will be answered by Viterbi algorithm.

Then we will compare 3 players' performance and test our results.