
Exercise Sheet 7

complete until Thursday, June 21st

Exercise

Estimate the amount of randomness involved in soccer play following these steps. First read all steps before starting to work on any of them.

1. (Model) Come up with a suitable probabilistic model of the outcome of a soccer game. Here is a suggestion. Model the outcome of a game by a number, by taking the difference between the goals scored by each time. For example, an outcome of 3:1 would correspond to +2, an outcome of 2:3 would correspond to -1. Assume each team has a strength and that the outcome of a game between two teams with strengths s_i and s_j is determined as follows: with some probability p , the outcome follows a normal distribution with mean $s_i - s_j$ and some variance σ . With probability $1 - p$ the outcome follows a normal distribution with mean 0 and some variance τ . Note that σ , τ , and p are independent of which teams play.

Feel free to come up with any other model if you don't like this one. Make sure though that it contains at least one hidden variable ;-)

Use the Wiki to discuss about the sense or nonsense of various models.

2. (EM algorithm) Formulate the EM algorithm to compute those parameters that maximize the likelihood of some given data with respect to your model from Step 1. In the suggestion above, parameters would be the strength of each team, the variances σ and τ , and p . First assume that your data is just the scores from n different games of the same two teams. Then assume you have data from m teams, with n games between each pair of teams.

3. (Data) See that you get some real data. Select at least three different teams with at least 10 games between each pair. As soon as you have some nice data, make it available to the others via the Wiki. The more data (in particular, the more games per pair of teams), the better.

4. (Implementation) Implement your algorithm from Step 2, run it on your data from Step 3, and output the estimated parameters of your model from Step 1. Make sure that you have tested your algorithm on some extreme data, and that it produces the results one would expect. For example, if one team has consistently won against another, its estimated strength should be larger.

5. (Discussion) What have you learned from your results? How much randomness is involved in soccer play? Post a summary of your results on the Wiki (following the instructions given there).