4F13 Machine Learning: Coursework #2: Probabilistic Ranking

Carl Edward Rasmussen & Zoubin Ghahramani

Due: 4pm February 24th, 2015 to Laura Reed, room BNO-37

In this assignment, you'll be using the (binary) results of the 2011 ATP men's tennis singles for 107 players in a total of 1801 games (which these players played against each other in the 2011 season), to compute probabilistic rankings of the skills of these players.

The match data is provided in the file `tennis_data.mat`, which contains two matrices: `W` is a cell array, whose `i`'th entry is the name of player `i`, and `G` is a 1801 by 2 matrix of the played games, one row per game: the first column is the identity of the player who won the game, and the second column contains the identity of the player who lost. Note, that this convention means that variable y_g (the game outcome) in the lecture notes is always +1, and can consequently be ignored. Some rows will appear more than once (corresponding to two players having played each other several times with the same outcome).

Your answers should contain an explanation of what you do, and 2-4 central commands to achieve it (but complete listings are unnecessary). You must also give an interpretation of what the numerical values and graphs you provide *mean* – why are the results the way they are? Hand in a maximum of 5 pages.

a) 10% : Examine the data in `tennis_data.mat`. Compute a ranking for each player based simply on the empirical ratio of number of wins to total number of games played. You may find the bar plot in `cw2.m` useful. Is this a good way to estimate player skills? Why/why not?

b) 10% : Complete the code in `gibbsrank.m`, by adding the lines required to sample from the conditional distributions needed for Gibbs sampling for the ranking model discussed in lecture 6 and 7.

c) 10% : Run the Gibbs sampler for 100 iterations. Plot some of the sampled player skills as a function of the Gibbs iteration. Does it look as though the Gibbs sampler is able to move around the posterior distribution? How far do samples have to be spaced apart, to be roughly independent? Explain your answer.

d) 10% : Explain what is meant by *convergence* of a Markov chain. What type of object does it converge to? Examine whether Gibbs sampling seems to converge? Do you reach the same distribution from runs with different pseudo random number seeds?

e) 10% : Generate 100 roughly independent samples from the Gibbs sampler, by saving only samples spaced by the distance you estimated in question c). Based on these samples, compute a ranking of the players, by computing the average probability that each player will win against another (randomly chosen) player. The Φ function is implemented in matlab as `normcdf`. Compare this ranking to the one obtained in question a). Explain the differences (if any).

f) 10% : Based on the samples generated in question e), find the probability that each of the 4 top players according to the ATP rankings in the handouts for lecture 6 would have of winning against each other (a 4 by 4 table may be a convenient way to represent this).

g) 10% : Do inference in the model instead, by running message passing and EP using `eprank.m`. Examine how many iterations are necessary for this approximate inference scheme to converge? Explain how you judge convergence?

h) 10% : Produce a ranking based on the results of question g), after running sufficient iterations for convergence. How does this ranking differ from the Gibbs ranking from e) and the empirical ranking from a). Explain your findings. Which ranking do you think is best?

i) 10% : Replicate question f) for message passing/EP? Comment on the differences, if any.

j) 10% : Repeat f) and i), but this time, report the probability that the player has a higher skill (not the probability that he would win a match). Comment on the difference to the results from f) and i).