L5: learning latent rankings from pairwise comparisons

Caterina De Bacco

May 1, 2025

1. Homework

a. Determine what is the best scoring system between SR and BT (and others if you want to add more), in the soccer datasets in the notebook. Clearly define what metric you consider and how do you make sure that each model is properly tuned.

b. Is the Spanish La Liga getting narrower and narrower? Is it becoming less predictable? Answer to these question by using the datasets from various years and measuring statistics as depth of competition, β, performance metrics.

c. The dataset contains additional detailed information of events happening during the game. Can you use them to improve prediction performance? Get some main statistics (e.g. number of shots, passes, etc...) from the event data and apply an arbitrary classifier to predict outcomes. Compare to what you obtained by learning scores with no additional information.