Predicting Betting Odds in European Club Football

With Bayesian Inference

Sebastian Adam, 01525543

Seminar in Artificial Intelligence, 194.103

Table of contents

1. Problem Description & Goals
2. Observations
3. Probabilistic Program
4. Results & Evaluation
5. Lessons Learned

Problem Description & Goals

Goals

- **Main goal** is to predict betting odds of yet to be played football games
- (+Bonus) goal is to get an understanding as to which parameters influence the betting odds the most.
  Examples for these parameters include:
    - Goals in the last **x** amount of games
    - Wins in the last **x** amount of games
    - Average rating of players on the field
    - etc.

Observations

Data

Observations are taken from a Kaggle database [2] which includes the following key metrics:

- More than 25.000 of lead championship matches of 11 European Countries
- Data from seasons 2008 to 2016
- Player attributes sourced EA Sport's FIFA of over 10.000 players
- Match data that includes team lineups and betting odds from up to 10 providers

Data Preparation

Major part of this project was to prepare the data for the actual inference. This data preparation resulted in the following parameters:

- Average points in the last x games
- Average rating of players on the field
- Average points against enemy
- (Historical average betting odd against enemy)
- etc.

Probabilistic Program

Idea

Two Approaches:

- Linear Bayesian Regression
  - - Not very good results
  - + Still useful to inspect parameter weights
- Bayesian Neural Network
  - + Far better results
  - - Weights do not represent meaningfulness of parameters anymore

Implementation

Linear Bayesian Regression (LBR) from pyro.ai [1].

- Simple linear regression
- Normal Distribution as Priors
- Stochastic variational inference optimizing the Evidence Lower Bound (ELBO)

Bayesian Neural Network (BNN) from kaggle.com [3].

- BNN with two hidden layers
- Normal Distribution as Priors
- Stochastic variational inference optimizing the Evidence Lower Bound (ELBO)

Results & Evaluation

Results

	LBR		BNN	
	w/ bets	w/o bets	w/ bets	w/o bets
avg distance to closest	0.82	1.13	0.49	0.81
avg distance to average	1.10	1.47	0.71	1.12
within 0.5	0.32	0.25	0.61	0.47
within 1.0	0.63	0.49	0.81	0.68
within 1.5	0.82	0.67	0.88	0.78

Table 1: Results of the LBR and BNN

Parameter Importance

- w/ bets:
    - 15.55 - bet against avg
    - 9.01 - bet against last
    - 4.29 - rating enemy
    - 2.49 - home
    - ...
    - 0.14 - points itl against
- w/o bets:
    - 7.26 - rating enemy
    - 3.89 - points against avg
    - 2.60 - points avg
    - 2.01 - home
    - ...
    - 0.31 - rating

Lessons Learned

Learnings

- LBR is useful to get a sense as to which parameters are important
- Results of BNN exceeded the ones of LBR by far
- Betting odds of same matchups often are very similar, therefore historic results improve the model **a lot**

Issues Encountered

- Which parameters to choose?
  Choosing wrong parameters in some cases worsened the results
- Parameter importance
  BNN lead to far superior results, but comes at the cost of explainability
- How to choose lean rate and iterations?
- On which characteristics is the model evaluated?

Summary

The source code of the project can be found here:

https://github.com/S3basuchian/aisem_bets

References i

- Bayesian regression - introduction (part 1).
  `http://pyro.ai/examples/bayesian_regression.html`.
  Accessed: 24.02.2021.
- European soccer database.
  `https://www.kaggle.com/hugomathien/soccer`.
  Accessed: 24.02.2021.
- Simple bayesian neural network in pyro.
  `https://www.kaggle.com/carlossouza/simple-bayesian-neural-network-in-pyro/notebook`.
  Accessed: 24.02.2021.