Project 3: Competition

Football pass prediction
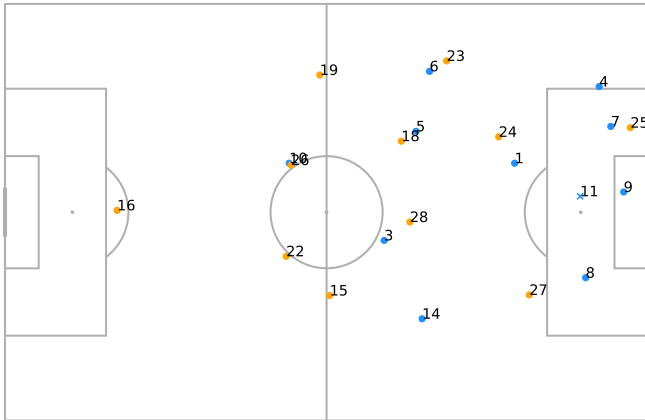
ELEN0062-1 | Introduction to Machine Learning

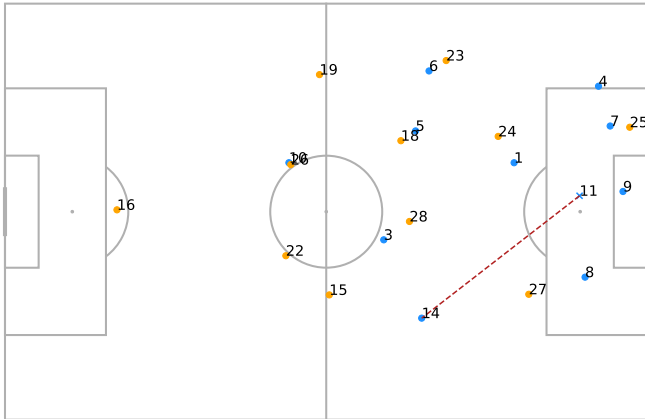**Goal:** Predict the next player who will receive the ball via a pass based on the position of all players and the ball at a given time.

Goal: Predict the next player who will receive the ball via a pass based on the position of all players and the ball at a given time.

About 12000 snapshots of the situation of the pitch when the ball is passed, 9000 for training and 3000 for testing.

The input variables are:

- the sender: the player who has the ball ($\in \{1, \ldots, 22\}$);
- the coordinates (x, y) in [cm] of all 22 players;
- the time in [ms] since the beginning of the concerned half-time period.

The output variable (to predict) is:

- the receiver: the player who receives the ball ($\in \{1, \ldots, 22\}$).
- only available for the training set samples.

Data correspond to 14 different games involving a Belgian football club during the 2014-2015 season.

Prediction format

We ask you to provide for the test set samples:

1. The predicted pass receiver ($\in \{1, \ldots, 22\}$)
2. Your estimation of the probability that each player is the pass receiver, i.e., 22 real numbers in [0.0, 1.0] that sum to 1.0.
3. An estimation of the accuracy of your predictions on the test set.

Evaluation metrics

Your predictions will be evaluated according to three scores:

1. Classification accuracy, Acc_{TS}

2. Brier score:
$$\frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} \sum_{j=1}^{22} (y_{i,j} - \hat{p}_{i,j})^2,$$

   with $y_{i,j} = 1$ if player j is the receiver for the ith pass of the test set, 0 otherwise and $\hat{p}_{i,j}$ is the predicted probability that the player j is the receiver for the ith pass.

3. Corrected accuracy:
$$Acc_{TS} - |\hat{Acc}_{TS} - Acc_{TS}|,$$

   Acc_{TS} is the actual accuracy on the test set and $\hat{Acc}_{TS}$ is your estimation of the accuracy.

The Competition

This project is organized as a *competition* using Gradescope.

During the competition: You can submit at most 3 times a day your predictions for the 3000 test samples. The first two scores, accuracy and Brier score, will be displayed computed on $\sim 50\%$ of the test set (= public test set).

After the competition: Once it is over, your final selected submission will be assessed on the other 50% of the test set (= private test set), using the three score metrics. The resulting *private* scores will be the final scores.

The examples appearing in the public/private parts of the test set will not be disclosed.

We expect more than a good score!

Shortly after the end of the competition, you are asked to submit a report which contains:

- A detailed description of all the approaches you have investigated.
- A detailed description of your approach to select and assess your model.
- A table summarising the results of your different approaches.
- All tables, figures and results should be analysed in-depth while avoiding unnecessary redundancies.
- Any complementary information that you want to analyse.

- 3 submissions per group per day.
- You can work in group of two or three students.
- Privately sharing code or data with other teams is not permitted.
- You can use any techniques or algorithms you want.
- You can use any library, but your results must be reproducible.
- You can not use external data, unless you get our approval first.
- Reference every external code, tool, algorithm, that you use (e.g., even Scikit-Learn).
- The use of AI is not prohibited, although not encouraged either, but you are expected to understand what you are doing and to follow the ULiège AI Charter.
- Obviously, do not use ready-made solutions to the problem.

All the resources will be available on the projects website.

$$https://iml.isach.be$$

A dedicated discussion forum will be open on Ecampus for your questions.

Questions?