Web Economics Course Project

Jun Wang junwang@cs.ucl.ac.uk
Emine Yilmaz e.yilmaz@cs.ucl.ac.uk

Submission Deadline: Midnight April 24th, 2016

This assignment is worth 20% of your total grade
Please submit assignment all required documents to Moodle

The goal of this project is to predict the user's click response to each auctioned ad impression in real-time bidding (RTB) display advertising. Specifically, given the information of the incoming bid request, the bid agent should estimate the probability that the user will click on its ad if it is displayed. For this purpose, you are asked to implement a machine learning regression model to predict the click-through rate based on the training data.

Kaggle Competition

We host a competition on Kaggle to make the evaluation standard and fair. You need to sign up the competition with your ucl.ac.uk email then join the competition via the invitation link:

https://inclass.kaggle.com/c/ucl-web-economics-algorithm-challenge-2016

The competition only allows UCL emails for registration. In addition, since this is a group project, you can build an up-to-3-people team in this competition, by clicking 'my team' in the competition webpage and inviting your team members. The team leader MUST send the account details of your team members with the following format to TA Bin Zou (b.zou@cs.ucl.ac.uk). NOTE: you are not considered as a valid participator of this competition if your information is not sent to Bin Zou (you will just be regarded as an observator of this competition).

Student Name	Kaggle Account	Kaggle Team
Bin Zou	Bin Zou	TA-baseline

iPinYou RTB Dataset

The dataset used in this project is sampled and reformulated from iPinYou RTB dataset and it can be directly downloaded from the kaggle competition webpage.

This dataset is in a record-per-line format. Each line contains the label, i.e., click (1) or no click (0), and the corresponding bid request & ad features, divided by '\t'. The detailed description of the data format is also given on the kaggle competition data page.

CTR Estimation

Normally, click-through rate (CTR) estimation is a regression problem where the label is a probability (of click). As such, logistic regression is a widely used model for CTR estimation. Besides logistic regression, you are encouraged to try other machine learning regression models.

The sample submission file can be downloaded with the train and test files. You should make sure that your submission file is a .csv. Besides the header line, it should have exactly the same line number with the test file. Each line gives the line ID in the test file and the predicted CTR. Any format error will lead the failure of submitting your prediction file.

Report Submission and Evaluation

Besides submitting your prediction file directly on kaggle, you are asked to upload a report describing:

- the statistical analysis of the training data and your feature engineering details;
- the details of your implemented machine learning model;
- some experimental analysis, such as the hyperparameters tuning etc.

The submission entry is given on the Moodle page.

The final score of this coursework will be based on the team ranking in the competition and the analysis presented in the report.

Contact

Besides Dr. Jun Wang and Dr. Emine Yilmaz, you are welcome to ask TA Bin Zou (b.zou@cs.ucl.ac.uk) for any questions.