Unity Data Science Homework

At Unity, we develop deep learning models for Real-Time Bidding (RTB)[1] at various ad exchanges. To bid for an ad impression, we estimate the optimal bid value using predicted install probability together with several other predictions. In this homework, your task is to train a model using the data sampled from our production environment and to predict the install probabilities for the ad impressions included in the test data. Note that the install probability predicted by your model will be used directly for estimating the optimal bid values of ad impressions. Therefore, it is important for the predictions to be as accurate as possible.

Here are the guidelines for the homework:

- Complete the homework with your language of choice.
- Use **ROC AUC**, **log loss**, and **prediction bias** to evaluate model performance. Feel free to use other metrics to discuss a model's merit.
- Perform exploratory data analysis.
- Keep code clean and organized.

Deliverables

You are asked to upload the following deliverables in a zip file:

a) A report (PDF or Jupyter notebook) detailing:

- Exploratory data analysis
- Modelling approach
- Performance evaluation of the model
- Explanation of design choices
- Discussion of future work

b) The source code used to produce the results

c) A CSV file contains the predicted install probability of each impression in the test data. The CSV file should have two columns: "ID" and "install_proba".

Data description

The training data includes the attributes and labels of each ad impression; The test data includes only ad impression attributes. The label is set to 1 if an ad impression leads to an install of the advertised game, otherwise the label is set to 0.

id:	impression id
timestamp:	time of the event in UTC
campaignId:	id of the advertising campaign (the game being advertised)
platform:	device platform
softwareVersion:	software version of the device
sourceGameId:	id of the publishing game (the game being played)
country:	country of user
startCount:	how many times user has started a campaign
viewCount:	how many times user has viewed a campaign
clickCount:	how many times user has clicked a campaign
installCount:	how many times user has installed games from this ad network
lastStart:	last time user started any campaign
startCount1d:	how many times user has started (any) campaigns within the last 24 hours
startCount7d:	how many times user has started (any) campaigns within the last 7 days
connectionType:	internet connection type
deviceType:	device model
install:	binary indicator if install was observed (install=1) or not (install=0) after impression

[1] Real-time bidding, Wikipedia: https://en.wikipedia.org/wiki/Real-time_bidding