CSCI 567: Mini-Project
Fall 2016

2016 BYTECUP Challenge

1 Introduction

In the mini-project, you will have the chance to explore an interesting machine learning problem by participating in the 2016 BYTECUP Challenge, which is being organized by IEEE China and Toutiao, and co-hosted by Tsinghua University.
The basic aim of the competition is as follows: Given a dataset with certain questions and a set of experts, you need to forecast which experts are more likely to answer which questions. Specifically, given a question and an expert, you need to calculate the probability of that expert answering the question. For more details, please refer to official competition website: `http://biendata.com/competition/bytecup2016/`.

2 Teams

- The project should be done in teams comprising of **up to 3 members**. Note that the competition officially allows teams of upto 5 members, but we are restricting the team size to a maximum of **3 members** for all teams from this class.
- Form a team of up to **3** students by **Sept 15, 11:59pm PST** and create your accounts on the official competition website. Make sure that all members of your team are registered under a single team name which begins with `MLCLASS_`. e.g. MLCLASS_awesome, etc.
- Fill out the sign-up form in this google spreadsheet with your team details. Please log in with your USC email.
- A team can have members from both sections of the class (30049D or 30095D).

3 Grading

- Your grade is bifurcated as follows: 1) your team's relative rank on the leaderboard **(60%)** and 2) the project report and code **(40%)**.
- Note that the leaderboard shows the ranking of all teams, not just those in this class. We will take the relative ranking among all teams into consideration. A better ranking will always lead to a better grade.
- Members of the same team will receive the same scores.
- The team that wins the 1st place among all teams in CSCI-567 class will get 5 bonus points in the total score for machine learning class. If the team wins the 1st place among all the teams on the leaderboard, it will get 5 more bonus points in the total score. We will consider the final ranking on the public leaderboard on **Nov 22** for grading and bonus points. We might announce an additional bonus policy later.

4 Deliverables

- Each team needs to write the project report in **NIPS format**. (6 pages maximum, including references; this page limit is strict). The NIPS LATEX format can be found here: `https://nips.cc/Conferences/2016/PaperInformation/StyleFiles`. Not following the format will lower your grade.

- In your report, you should cover the details of your solutions, including the general ideas, the way of data processing and cleaning, the learning algorithms and models you have tried, the results you obtained, and any other insightful thoughts during the competition. You should also describe how to run your code to get the results.
- All teams are advised to use Python as the programming language for their code. We also encourage you to use the Google Cloud Platform for running your codes, if needed.

5 Deadlines

- We will grade you based on your final released ranking on the online public leaderboard on **Nov 22**. Note that the official competition deadline is **Nov 20, 11:59pm UTC** (note **UTC**, not PST) and no changes are allowed after the deadline. Please make your final submissions on time.
- Each team should submit exactly one hard copy of the report to Locker 19 at PHE. The hard deadline for report submission is **Dec 12, 5pm PST** and any report submitted afterwards will not be accepted.
- Each team also needs to submit one soft copy of the project **report** and all the **code** via blackboard. This needs to be submitted in a single zip file titled CSCI567_Fall16_< *TeamName* >.zip (e.g. CSCI567_Fall16_MLCLASS_awesome.zip). The deadline for blackboard submission is **Dec 12, 11:59pm PST**.
- We will not accept any late submissions; neither for the hard copy nor for the online submission.

6 Policy on collaboration

In line with the rules of the competition, you are only allowed to collaborate within your own team. Your code will be analyzed to reproduce the results and compare its similarity to code from other teams. Any violation of the USC Integrity and Plagiarism policy will lead to an immediate "F" grade in this class and you might be subject to harsher penalties. So you are advised to strictly abide by the rules of collaboration for the competition.