EE379K
Spring 2016
Dr. Sriram Vishwanath

Final Project – Amazon Employee Access Competition (Kaggle)

Assigned: 3/22/16
Due: 5/6/16 by 11:59pm

For your final project you will all be participating in the Amazon Employee Access Competition on Kaggle, which is an online data science website that hosts machine learning competitions.

Step 1: Find a partner

For this project you will be working in groups of 2. It is upto you to find partners. I will create a Piazza thread to help people find partners.

Step 2: Get started with Kaggle

It is quite easy to get started with Kaggle. First, go to www.kaggle.com and create an account. Once you've done that, go here for details on the Amazon Employee Access competition: https://www.kaggle.com/c/amazon-employee-access-challenge

The challenge page has all the information you need, including the datasets, description of the challenge the data, as well as the expected output format. Read it carefully!

Step 3: Do a test submission

Download the sampleSubmission.csv file from the competition page and submit it on the submission page: https://www.kaggle.com/c/amazon-employee-access-challenge/submissions/attach

Kaggle will then tell you your “private score” which is just your accuracy on their test set, as well as your ranking. We only care about your **private**

score. We do NOT care about your ranking or your **public** score. For the sample submission, your private score should be 0.5

Step 4: Win

You have complete freedom in how you solve this challenge. You can use any machine learning algorithm you want, any training methodology you want, and any tools you want.

Submissions

You will submit the following on the due date:

1. A screen capture of your highest private score on this link: https://www.kaggle.com/c/amazon-employee-access-challenge/submissions
2. Your code along with instructions on how to run it.
3. Your submission CSV.
4. A two page report answering the following:

   a) How did your team go about tackling this problem?

   b) Which methods/algorithms did you try?

   c) What is your final methodology? Walk through it in detail, starting from data pre-processing. Explain all the machine learning algorithm(s) you used as well as the paramters you chose. Also discuss any external tools or libraries that you used.

Grading

Your base grade is composed of two parts: your report and your private score. The report will be worth 10 points, and your score will be worth the other 90.

- private score <= 0.8 : Grade = report score + private score * 88
- 0.8 < private score <= 0.85 : Grade = report score + private score * 93

- private score > 0.85 : Grade = report score + min(100, private score * 98)

Since this is a competition, bonus points will be given out for the following achievements:

- 20 bonus points for 1st place
- 10 bonus points for 2nd place
- 5 bonus points for 3rd place