Image by: Tang Yau Hoong

Poznań WiMLDS Workshop
TensorFlow on AWS

Part 4: Final Project

Pearson AI Products & Solutions
Poznań, 6 Feb 2019

GitHub repo

github.com/MateuszOtmianowski/wimlds_final_project

Final Project

Agenda

1. **Introduction**.
   Goal of the meeting, general rules, team creation (10 min)

2. **Teamwork** (90 min)

3. **Result presentation** (20 min)

4. **Summary** (5 min)

Goal of the project

1. Explore the dataset to understand it better.

2. Create a logistic regression model to predict student graduate admission success.

Questions you may consider in your EDA

1. Are there any missing values in the data set?

2. Does the average GRE score differ between universities with a different rating? What about TOEFL?

3. How research activity impacts the chance of being admitted? What might influence this?

4. Are GRE scores and TOEFL scores correlated with each other?

5. Which one is more difficult, GRE or TOEFL?

6. How university rating impacts the chance of admission?

   You can find explanation of variables [here](#).

Tips on logistic regression model

1. Use tensorflow to create logistic regression model.

2. You can train it on EC2 machine or locally.

3. It should predict whether a given graduate
    will be admitted or not.

4. Use whatever variables from the dataset you like.

5. Think how to evaluate your model.

Pro tips

1. Set up EC2 machine (no GPU needed) using *pearson_tensorflow_modelling_workshops* AMI or use a local environment with libraries we were using during earlier workshops.

2. Use tools you mastered during the previous workshops: Tensorflow, S3 and Athena usage are very welcome.

3. Time is limited, so team co-operation is crucial!

Pearson

Proposed timetable

You may (but don't have to) take it into consideration. Feel free to dive into a problem you think is especially interesting.

1. Configure your environment, find notebooks from the previous workshops, download dataset (10 min)
2. Create Athena table (5 min)
3. Explore dataset using Athena, find answer to the given questions (30 min)
4. Create logistic regression model in tf (35 min)
5. Prepare final presentation (10 min)
6. Present results (5-10 min per team)

Pearson

ALWAYS LEARNING