Machine Learning

Recitation 6, CMU 17-313, Fall 2020

Setup:

1. Ensure you've completed the setup instructions for today's recitation.
2. Pair up.
    a. If you're in the in-person recitation section, pair up with your neighbor and join the recitation section on Zoom. The instructor will assign you to a breakout room with your peer in order to facilitate screen sharing.
    b. If you're in the online recitation section, you will be randomly assigned to a breakout room by the instructor.
3. Join the #recitation-X channel on Slack, where X is the recitation section you are in.

Goal: During this recitation, student will:

1. have the opportunity to play with various machine learning frameworks and tools (e.g., such as pandas, LIME, and Jupyter Notebooks.)

Context: In lecture, we talked about using the Titantic dataset to make predictions on whether or not passengers would survive given features in the dataset. We saw how gender was one feature that predicted if a passenger would survive, but during class several other ideas were proposed as well. For example, one might consider if fare paid was a good proxy for predicting if a passenger would survive.

Part 1 (35 mins): During this recitation, within your group, you should identify a question to ask about the data set. As we stated before, one example (do not choose this one) is whether the fare paid was a good proxy for determining whether a passenger would survive. Once you identify your question, train your model, and use LIME to look at both a a.) lucky passenger and an b.) unlucky passenger in the dataset and determine whether you were correct or not.

Using a link provided by the instructor, construct a single slide that shows the following:

1. the question you asked.
2. the features you selected.
3. screenshots of the LIME explanation for both the unlucky and lucky passenger.
4. form an initial conclusion, does your hypothesis hold? (at least for these two examples and obviously not generally – that is out of scope for a 50 minute recitation, as we're only trying to have an opportunity to learn the tools and not generate new insights about the dataset, not understand the intricacies of machine learning.)

You should prepare a 1-2 minute presentation and be ready to present it to the class.

Part 2 (15 mins): Present your findings to the class.