MACHINE LEARNING 2- PROJECT PROPOSAL

Selvyn Perez, Ritika Agarwal, Tanaya Pole

- Yelp users upload around 100,000 photos a day to a collection of tens of millions, and that rate continues to grow. In fact, the growth rate for photos is outpacing the rate of reviews. These photos provide a rich tapestry of information about the content and quality of local businesses.
- We decided to take up the Yelp Dataset challenge as our final project. We agreed that this dataset can be a good learning experience and the applications and use of this research could be applied in multiple different areas. We wanted to apply classification techniques we learned through coursework.
- The data consists of business data, user information and photo data including the caption and classification. We are going to focus on using the photo data for image classification. There are more than 2 million images which is large enough to train a deep network.
- We will use Convolutional Neural Network and Feedforward Neural Network and compare the results. As of now, we have decided to use Torch as the framework and pytorch the library. In order to obtain sufficient background to apply these networks we will be using the material provided during the duration of our course and multiple resources which are available online.
- The performance of our model will be judged by calculating the accuracy of predictions based on a set of pre-classified images. We will also compare our classifier performance to other open image recognition network performances.
- Rough schedule of project completion:
    - Mid November - Complete preprocessing of data set

- End November - Complete initial training model
- Mid December - Complete presentation, and final analysis