Multiclass classification using Neural Networks

Akshay Kumar Varanasi

1. Your project problem, question, or ask. What is the problem you are interesting in looking at in your term project?

The problem statement which I am interested for final project is multi class classification using Neural Networks on large datasets. In the field of Machine Learning, Neural Networks is hot topic due to promising results in various fields like Computer Vision, NLP etc. So I would like to learn it through this project for image classification. Since the datasets which I will be dealing is huge, it is better to create a database using PostgreSQL. So in this way, I will be learning both Neural Networks and SQL.

2. Datasets used. What datasets are you planning on using in your project and how will they help you meet your objective in part 1?

Dataset that will be used is Kaggle-Inclusive Image Challenge
https://www.kaggle.com/c/inclusive-images-challenge/data
This is Kaggle challenge by Google AI to develop models that are robust to blind spots that might exist in a data set, and to create image recognition systems that can perform well on test images drawn from different geographic distributions than the ones they were trained on.
This is basically image classification problem with multiple classes and since data set is huge I have no problem training the Neural Network to do the classification and also use SQL in the process.

3. What tools do you plan to use to clean, analyze, store, and/or visualize your data?

Since the database is too big for excel I will be using PostgreSQL for storing and for classification/analyzing/visualizing R will be used as it has packages like “neural net” to do classification using Neural Nets.

4. What do you expect the results, products, lessons learned, and/or deliverables to be?

Things I will be learning is how to do multiclass classification(training and testing), how to use Neural Networks. Since I am working with real world data which is really really large (stored in databases generally), I will be learning how to create Databases for the dataset and read from them after doing necessary pre-processing like cleaning in SQL. Since the main analysis is done in R, I need to read the data in database using R and do classification, which also useful to learn(how to use both together). Through this project, I will also learn how to use TACC resources because in industries we need to use high computing resources for tackling their problem. All the analysis and lessons learned will be submitted as a report along with the codes. The result of the project is classification result of the images obtained using Neural network model.