Large Scale Machine Learning - Assignment 3
Abhishek Sinha, Arun Sai, Ashish Bora

March 28, 2016

1   Q1

- Model: $\ell 1$ regularized logsitic regression

- Private Score: 0.896

For this question we generated additional features from the original categorical features. All pairs and triples of the original categorical features were generated. Also used one-hot encoding.

2   Q2: XGBoost

- Model: Boosted Decision Trees.

- Private Score: 0.883

All pairs and triples of the original categorical features and their frequencies (i.e, number of times a particular pair or a triple occurred) were used to train the model. Doesn't use one-hot encoding. Best parameters found using cross validation: $learning\_rate = 0.2, n\_estimators = 100, colsample\_bytree = 0.1, max\_depth = 6$

3   Q2: XGBoost with one-hot encoding

4   Q3

Random Forests..
Ensemble..