John Kim

MATH 4630

Professor Mark Ellingham

April 18, 2017

**Rental Interest Prediction**

**Overview**

1. View summaries/statistics of correlated data (statistic vs. statistic)
   1. Maybe already available on Kaggle?
   2. Most correlated features
2. Create classifier based on highly correlated statistics
   1. Translate MATLAB classifier into Python
      1. Using Ex2 and Titanic as reference

**Steps**

1. Download/install Docker
2. Implement containers from Kaggle
   1. Requires VirtualBox
   2. Setup Kaggle kernels
3. Use containers to analyze data

**Main**

1. Data pre-processing
   1. Read JSON file
   2. Translate JSON into pandas then into numpy matrix
   3. Print some statistics
   4. Extract attributes
   5. Add bias unit
   6. Translate date to usable attribute
   7. Translate target to numerical value
   8. Normalize values
   9. Split training set into train and cross-validation set
2. MATLAB translation
3. Feedback
   1. Use precision & recall