Zach Fontenot & Zack Vickers

Machine Learning

Dr. Kawsar

Machine Learning Project: Phase 1 Presentation Points

*Platform*

* Python
  + Open source libraries
  + Good documentation
* IDE
  + Atom: Simple user interface
  + PyCharm: JetBrains application -> Familiar, good functionality
* Communication
  + Github

*Goals*

* Modular design
  + The application should have set modules with easy to understand input and outputs
  + Low coupling
* Versatile
  + Should be able to take a large amount of inputs and output into a desired format

*User interface*

* TKinter
  + Open source command-line interface that packages user selections into JSON format
* The application can be ran in any directory once installed
* Error checking
  + Some menu options are mutually exclusive, entering two of these will cause the program to exit
  + Some menu options are set a default, so choosing no options for a category will toggle the default

*Tools*

* Python Libraries
  + Statistics– Math library for python

*Data Processing*

* Intervals
  + An interval object kept track of all data for a set interval
  + Interval length could be determined by the user
    - Determined by number of lines rather than time
  + Left and right foot kept separate
    - Left and right foot statistics were similar. These two sets were kept separate to effectively double the amount of training examples
    - This can be changed later if results are not ideal
* Splitting Dataset
  + The user was given the option to split data into learning, testing, and degree data sets with variable proportions
* Feature Generation
  + The statistics class was used to generate mean, median, variance, and standard deviation for all interval objects.
  + More features can be added later with minimal refactoring
* Cleaning Data – WIP
  + Removed empty lines
  + Some key features of bad data were identified
    - Data at the beginning and end of the files were likely to be bad data
    - Outliers determined by the interquartile range could be removed
    - High variance over a short time period is an indicator of bad data
  + Depending on what activity is preformed, different cleaning measures may be needed