Laboratory Exercise - 3

Objective

- This case study provides an opportunity to demonstrate our ability to combine data sets and produce meaningful analysis. Specifically, we would like to provide a decision maker with more than just data – we want to provide insights, understanding, and wisdom. This exercise allows the student an opportunity to demonstrate progress (or mastery) of learning objectives 1, 2, 3, 4, and 5:

1) **O**btain data and understand data structures and data elements.
2) **S**crub data using scripting methods, to include debugging, for data manipulation in R and other tools.
3) **E**xplore data using essential qualitative analysis techniques including descriptive statistics.
4) **M**odel relationships between data using the appropriate analytical methodologies matched to the information and the needs of clients and users.
5) I**N**terpret the data, model, analysis, and findings. Communicate the results in a meaningful way.

Instructions

- The research question is can we use algorithms and compute to identify clothing items? Specifically, can we determine which algorithm and compute methodology provides us the most efficient approach for classifying simple fashion images?
- Using the base samples available from Zalando Research:
    - https://github.com/zalandoresearch/fashion-mnist
    - Review the data – clean as appropriate
    - Provide an initial data analysis
- Implement at least two approaches for classifying the digits – examples below:
    - Naïve bayes
    - Neural Networks
    - Keras
    - Azure ML
    - IBM DSX
    - Boosted trees
    - Linear classification
    - Your choice
- Answer the following questions:
    - What is the accuracy of each method?
    - What are the trade-offs of each approach?
    - What is the compute performance of each approach?

Additional Instructions

- Don’t forget what you learned in your previous courses; do your own work, document any assistance, use comments for clarity.
- Report results for both the training set and the test set.
- Feel free to use any software to conduct your analysis and produce your graphics

Submission Items

- Report with graphics
- Supporting notebook for the report
- Description of your tool and methodology if you used another statistical/software package