CSC 466: Project Proposal

Team Members: Tatjana Ellis, Younis Ereiqat, Jason Ku

Summary Statement

Our objective is to create tutorials to introduce unsupervised learning concepts to people who are unfamiliar with the topic. Students with minimal to no experience with unsupervised learning ought to be able to start with these tutorials and gain a basic working knowledge of unsupervised learning and how to implement a model to solve a relevant problem.

Specific Aims

- Make a comprehensive tutorial to introduce people new to ML:
    - What is unsupervised learning?
    - When to use (vs. supervised)?
    - What are the types of unsupervised learning problems?
    - How to create an unsupervised ML model? (In steps)
        - Types of models (how to choose?)
        - Go into detail with a common one
        - First: Easy, well-correlated model example
        - Next: NBA Dataset (Predicting winning team; use association model)
            - Show how to tweak features used to train
            - Show real-life data, and how it ofen doesn't perfectly correlate/requires a lot of tweaking

Roles

Tatjana Ellis - Data Cleanup, ML Engineer, & Tutorial designer

- Organize and clean up dataset(s)
- Develop the ML models
- Design tutorial (layout, ..) & help implement the examples

Younis Ereiqat - Data Acquisition, ML Engineer, & Tutorial Implementer & designer

- Find the best dataset(s) to use
- Develop ML models
- Help implement & design the tutorial example

Jason Ku - Data Cleanup, ML Engineer, & Tutorial Implementer

- Organize and clean up dataset(s)
- Develop ML models
- Help write tutorial for simple example

Final Deliverable

The final deliverable will be a series of interactive Python notebooks (3 .ipynb notebooks), including visualization of models and their results, accessible through GitHub and Google Colab.

Timeline

Week 3

> Finalize project details
> Set tasks to start working on

Weeks 4 & 5

> Find a well-correlated dataset for the simple, first example
> Find a large dataset to run an association model on (suggestion: NBA dataset)
> Clean up & organize datasets
> Write a Google Colab notebook to introduce the concepts surrounding unsupervised learning (Lesson 1)

Week 6

> Write a Google Colab notebook to run through a simple, straightforward unsupervised learning example (using simple dataset; Lesson 2)
> Create and train association model for NBA dataset that the student should use as an example
    > First find association rules between individual features
    > Then find association rules between groups of features

Week 7 & 8

> Write Google Colab notebook to guide student through creating their own unsupervised learning model using the NBA dataset (As described in Week 6; Lesson 3)
    > Emphasis on real-life data: Often needs a lot of tweaking, can be difficult to find good correlations, …
    > Think of good questions for the student to answer/learn from

Week 9 & 10

> Finish up notebooks; final edits and revisions