IBM

Contents

This directory is broken down to showcase the following projects, reports, and skillsets.

Deep Learning (Python)

This project was a team effort focused on using deep learning on ~50,000 sattelite images and temperature data accross multiple years to predict crop yield. Ground truth data consisted of six years of crop outputs for multiple districts in India. We used this data to train many CNN and LSTM models of different architectures on a Google Cloud GPU.

The **Deep Learning** directory consists of the:

1. project poster
2. report
3. highlight of the project on social media.

Language used: `Python`

Big Data Mining (Spark)

Big data is often used in sales and market contexts to provide data driven insights which add value. This folder consists of two big data algorithms.

- The first is market basket analysis which uses the A-priori algorithm to understand purchase behavior of consumers. Such insights can aid in promotional activities, marketing, cross selling, store design, etc.... The algorithm is able to find products frequently viewed together and in this case is used as an example to power a recommendation engine on an online website.The browsing session for each of the 31000 customers is stored on each row of the data file.
- The second is implementation of a machine learning model, support vector machines, on Spark using a variety of gradient descent algorithms. Mini batch gradient, stochastic gradient, and batch gradient were tested to examine differing convergence times.

The **Big Data Mining** directory consists of the:

1. apriori.py
2. svm.py

Language used: `Python`, `Spark`

Machine Learning (R)

Machine learning can also be performed using R. In this project, the task was to create a model for daily forecast of traffic in a parking lot for the next month using historical data. In this exercise, I demonstrate my reasoning on simple model building of R, ranging from topics on how to evaluate feature importance and performance of cross validation.

The **Machine Learning** directory consists of the:

1. Car traffic report

Language used: `R`

Data Wrangling, Exploratary Data Analysis, Tidyverse (R)

In this directory, I created a exercise meant to teach indivudals the tidyversre, visualization, data wrangling, and scoped verbs. This can be seen as investigative journalism where I examine the impacts of redlining (discriminatory housing practices) to populations

The **Tidyverse** directory consists of the:

1. The impacts of redlining report

Language used: `R`