Satej Soman
CAPP30254: Machine Learning for Public Policy
Spring 2019

HW 1
DIAGNOSTIC ASSIGNMENT

1 Data Acquisition & Analysis

1.1 Chicago Open Data Portal

Chicago crime data is available, filtered by year, from the Chicago Data Portal (`https://data.cityofchicago.org/browse?category=Public%20Safety`). We can download this data and load it into a Pandas `DataFrame`:

```
from pathlib import Path

import pandas as pd
import requests

# download crime data if we don't have it locally
base_url = "https://data.cityofchicago.org/api/views/{}/rows.csv?accessType=DOWNLOAD"
crime_resources = {
    2017: (Path("./crime_data_2017.csv"), "3i3m-jwuy"),
    2018: (Path("./crime_data_2018.csv"), "d62x-nvdr"),
}

for (year, (path, identifier)) in crime_resources.items():
    if not path.exists():
        url = base_url.format(identifier)
        print("{} data not found locally, downloading from {}".format(year, url))
        response = requests.get(url)
        with path.open("wb") as f:
            f.write(response.content)

crime = pd.concat([
    pd.read_csv(crime_resources[2017][0]),
    pd.read_csv(crime_resources[2018][0])
])
```

1.2 Summary Statistics for Crime Report Data, 2017-2018

2 Data Augmentation & APIs

2.1 Chicago Crime Reports, Augmented with ACS Demographic Information

3 Analysis & Communication

3.1 Changes in Crime, 2017-2018

3.2 Analysis of Jacob Ringer's Claims

3.3 Key Findings

3.4 Caveats & Limitations

4 Probability Exercise

a)

b)

c)