

Task 1

Signal Processing and Machine Learning

Machine Learning



PROBLEM STATEMENT:

The Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system has lost the locations of some of the crimes in Chicago. They need an approximate location of the crime from the remaining data. The dataset is available with the police department and it reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2012 to 2017.

Task:

For the lost locations the related data to those locations is still with Police Department. You are given the data related to lost locations of the city. Your task is to predict the exact location of the Crime by the help Neural Networks algorithm.

Dataset contains following features:

ID - Unique identifier for the record.

Case Number - The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.

Date - Date when the incident occurred. this is sometimes a best estimate.

Block - The partially redacted address where the incident occurred, placing it on the same block as the actual address.

IUCR - The Illinois Uniform Crime Reporting code. This is directly linked to the Primary Type and Description.

Primary Type - The primary description of the IUCR code.

Description - The secondary description of the IUCR code, a subcategory of the primary description.

Location Description - Description of the location where the incident occurred.

Arrest - Indicates whether an arrest was made.

Domestic - Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.

Beat - Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts.

District - Indicates the police district where the incident occurred.

Ward - The ward (City Council district) where the incident occurred.

Community Area - Indicates the community area where the incident occurred. Chicago has 77 community areas.

FBI Code - Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).

X Coordinate - The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.

Y Coordinate - The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.

Year - Year the incident occurred.

Updated On - Date and time the record was last updated.

Latitude - The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.

Longitude - The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.

Location - The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

Dataset Link:

<https://drive.google.com/open?id=12JgRDsv460rR5atQw2Bha0M0UGSd99lp>

It contains training data as well as validation data.

GUIDELINES:

Full implementation should be done in Python and only computation libraries NumPy and Pandas should be used. High level Machine learning frameworks or libraries like Scikit-learn, PyTorch, Keras, Tensorflow etc. should **NOT** be used.

EVALUATION METRICS:

- Accuracy on a test set [which will be provided later]
- Features used
- Algorithm Implementation [Neural Networks]
- Loss function used

COMPONENTS REQUIRED:

Laptop or Desktop with Python installed with NumPy and Pandas libraries.

SUBMISSION:

- A folder is to be made, name of the folder should be **rollno_domain/subdomain_taskname** . This folder will contain:
 - Codes: All the codes files along with any external library files required to execute it.
 - Trained weights: After training the Neural Network, the weights should be saved in **.csv** (Comma-separated Values) format.
- After making the folder, compress the folder to **.zip** format and upload to your **Google Drive**, then get the sharing link of the folder from Google Drive and submit that link to induction portal.

RESOURCES:

- [Linear Regression and Machine Learning](#) [Subset of Machine course by Andrew Ng which contains videos for learning concepts required for your task]
- [Original Andrew Ng's Playlist](#)
- To use NumPy and Pandas follow their documentations for your python version