SEQOME-DL: An Integrated toolkit for rapid development and optimization of Deep Learning models

Quick Start Guide

1. INPUT FORMATS-
   * A feature file with the first “column” containing id’s and the rest “columns” containing features.
   * A meta/output file with the first column containing id’s and the second column containing the output values. (Care to be taken to One-Hot-Encode the outputs if output is not numerical)
2. HOW TO USE TRAIN SCRIPT-
   * AUTO MODE-
     + It creates the most optimal network architecture and trains it using the input files using a 75%-25% split.
     + The user gives the feature and output file in command line.
     + Command Line Syntax for execution – **“python s\_create\_n\_train\_auto.py Feature\_file\_path Output\_file\_path“**

**Eg**. python s\_create\_n\_train\_auto.py C:\Users\Desktop\diabetes\_with\_patient\_ID\_2.csv C:\Users\Desktop\diabetes\_with\_patient\_ID\_meta.csv

* + - The user is shown a network summary and model image of the architecture.
  + MANUAL MODE-
    - It creates the network according to the desire of the user.
    - The user gives the feature and output file in command line and with the help of prompts the user can create the desired network architecture. (To be noted that the answer to the prompts are Case Sensitive and should be answered in the same way as the prompt suggests)
    - The user can add any number of layers he wants, with any number of nodes. The activation functions currently supported are relu, sigmoid and tanh.
    - Command Line Syntax for execution – **“python s\_create\_n\_train\_man.py Feature\_file\_path Output\_file\_path“**

**Eg**. python s\_create\_n\_train\_man.py C:\Users\Desktop\diabetes\_with\_patient\_ID\_2.csv C:\Users\Desktop\diabetes\_with\_patient\_ID\_meta.csv

* + - The user is shown a network summary and model image of the architecture.

1. HOW TO USE TEST SCRIPT-
   * The test script is to be used to test new data based on the model architecture created.
   * The user gives the feature and output test file in command line.
   * The user is given an accuracy score which show the accuracy of the model.
     + Command Line Syntax for execution – **“python s\_test.py Feature\_file\_path Output\_file\_path“**

**Eg**. python s\_test.py C:\Users\Desktop\diabetes\_with\_patient\_ID\_2.csv C:\Users\Desktop\diabetes\_with\_patient\_ID\_meta.csv

1. HOW TO USE FEATURE SELECTION SCRIPT-
   * This script takes feature and meta/output file as input from the command line.
   * It gives various statistical and visual representation of different feature selection methods.
   * It gives the accuracy, ROC-AUC score of the model and uses perturbation mechanism to rank the features. P-values using ANOVA test and correlation heat map along with correlation pair plot is also shown as output.
     + Command Line Syntax for execution – **“python s\_find\_feature.py Feature\_file\_path Output\_file\_path“**

**Eg**. python s\_find\_feature.py C:\Users\Desktop\diabetes\_with\_patient\_ID\_2.csv C:\Users\Desktop\diabetes\_with\_patient\_ID\_meta.csv