

Network-based Protein Structural Classification

Mahboobeh Ghalehnovi

June 2019

1 Data processing and Feature Extraction

This section is done by Khalique Newaz and I do not have any script for this part.

2 Classification

Required Python packages for running all scripts of this section is scikit-learn 0.19.2, tensorflow 1.8.0, and skimage 0.14.1. The version of Python is 2.7.15.

2.1 Logistic regression-based classifier

1. For running Logistic regression-based classifier for individual features for a given dataset first go to the folder named “LR_individual” and then run “python RunLR.py”. (Note before running the script first the name of dataset should be set (let’s modify line 28 in the script “RunLR.py” for the desired dataset and then run the file).
2. For running Logistic regression-based classifier for integrated features first go to the folder named “LR_integrated” and then run “python RunLR_PRS_36dataset.py”.

2.2 Deep learning-based classifier

For running Deep learning-based classifier first go to the folder named “DL” and then run “python RunDL.py”. The script needs to be run for each of the data sets separately and can be run on CPU or GPU. (Note before running the script first the name of dataset should be set (let’s modify line 13 in the script “RunDL.py” for the desired dataset and then run the file).