BINF 610 Applied Machine Learning (Spring 2023)

HW4

Due: May 12

DeepNet

In this assignment, you are asked to train deep neural net to predict 10 subcellular localizations for proteins.

You will use the same data from HW3, and each protein is represented by a 400-dimension vector of dipeptide frequency.

Evaluate the performance from 5-fold cross validation using the confusion matrix as described in HW3.

1. Train and test a fully connected feedforward NN with three layers:
   Input layer 400,
   Hidden layer 200 (ReLU)
   Output layer 10 (softmax)

2. Train and test a fully connected feedforward DNN with 6 layers: 400, 200, 100, 50, 25, 10.

   a) Use sigmoid activation function, and random initialization using a normal distribution with a mean of 0 and a standard deviation of 1.

   b) Use LeakyLU activation function, HE initialization, and Adam Optimization

Discuss and compare performance of the various settings.