NTUST, CSIE
Machine Learning (CS5087701), Fall 2025

Homework 3 (12pts)

Due date: Nov. 18

Question 3.1. [12pts] Consider the Adult (A) dataset (https://archive.ics.uci.edu/ml/datasets/Adult) from UCI (http://archive.ics.uci.edu/ml/) and the MNIST dataset used by LeNet in 1989 (MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges and follow the guidelines below to obtain your results.

  (a) Let us use either shallow or deep networks as the modeling method to solve the problem. You may consider networks with only one to three hidden layers as shallow networks and others as deep networks.

  (b) Find a network with about the right amount of hidden units to produce effective results. How many hidden layers could be considered enough? You may consider a shallow network that is good enough to obtain excellent result. What could be the *minimum* number of hidden layers that we need in this case? On the other hand, you can also attempt to use a deep network as the model. How deep can you choose so that you do not run into the overfitting situation?

  (c) Do you observe the *vanishing gradient* phenomenon in your training? You need show some evidence to confirm your observation.

  (d) Can we have a different or better learning result if we explain the output unit as probability, meaning a logistic regression kind of network is discussed here?

  (e) Are the performance and the networks that we used different the two datasets? What makes them to be different? Can you explain this?

  (f) There are packages to use. For instance, you can either use the functions from Weka 3 - Data Mining with Open Source Machine Learning Software in Java (waikato.ac.nz), or from scikit-learn of Python. Using Matlab or R and writing code on your own is also ok, but not necessary.

  (g) After your experiments, you should give all necessary details of your study in a mini report of around two to four pages with a discussion section. In your report, some of the following items are recommended to include:

    i. All the experimental settings and meta-parameters used for the methods.

    ii. The prediction accuracy with cross-validation and possible different data partitions.

After all, the key to receive high score relies on how you can explain your result well, such as why you have good or bad result, for different settings or for different datasets. Also, bonus goes to any creative ideas!