Midterm
MGT-302

Due to 05.05.2021

Please submit your answers by the end of 05.05.2021 to Teng Andrea Xu: andrea.xu@epfl.ch. For the implementations, you are free to choose your favorite programming language.

Problem 1. [50]
Implement a Neural network with two hidden layers to do classification on a small dataset P1.data, which contain $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$ and the labels are binary ($y_i \in \{0, 1\}$). You need to use the first 400 instances as your training data and the remaining data as your test data. Train your neural network and compute your test error. You need to show the convergency of your algorithm and report your test error.
In your implementation, set $h_1 = 8$ and $h_2 = 16$, where h_i is the number of neurons in i-th hidden layer and use Sigmoid as your activation function. Use the logistic regression log likelihood as your loss function:

$$\sum_{i=1}^{m} (y_i \log f(\mathbf{x}_i) + (1 - y_i) \log(1 - f(\mathbf{x}_i))).$$

You are free to select the unspecified parameters, e.g., the learning rate, the optimizer, etc.

Problem 2. [50]
You are familiar with the concept of structure learning. In this problem, you will learn about a new structure learning algorithm called Chow-Liu. The goal of this algorithm is to learn a graph with tree structure that can best describe the dependencies within a dataset. Below, you can see an overview of the algorithm to find the tree structure within p variables.

1. Start with a complete graph of p nodes.

2. Compute weights (mutual information) of each (possible) edge

3. Find a maximum weight spanning tree (MST): This is a tree with p nodes and the greatest total weight.

- Let $(X, Y) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ be a pair of Normal random variables with mean μ and covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{2\times 2}$, the mutual information between them is given by

$$I(X;Y) = -\frac{1}{2}\log\left(1 - \frac{\mathbf{\Sigma}_{12}^2}{\mathbf{\Sigma}_{11}\cdot\mathbf{\Sigma}_{22}}\right).$$

Implement an algorithm `Mutual`(S) that estimates the mutual information between two Normal random variables X and Y given a set of samples, i.e., $S = \{(x_i, y_i)\}_{i=1}^n$. To evaluate your algorithm, generate a set of i.i.d. samples with size n from

$$\mathcal{N}\left((0,0), \begin{bmatrix} 1 & 0.3 \\ 0.3 & 2 \end{bmatrix}\right),$$

and report the estimation error of your algorithm with respect to the sample size n ($n \in [10^3, 10^5]$).
- Implement the Chow-Liu algorithm, `ChowLiu`$(\mathbf{W})$, that takes an $p \times p$ matrix $\mathbf{W}$ as input, and returns the adjacency matrix of the final tree. $\mathbf{W}_{ij}$ is the weight of the edge between nodes i and j.