Assignment Late Submission Policy: There will be a 20% penalty on the assignment mark for late submissions submitted within 24 hours after the deadline, and a 50% penalty for late submissions submitted between 24-48 hours after the deadline. Assignments submitted after 48 hours after the deadline will not be accepted.

1. (**30 marks**) We are to train an artificial neural network (ANN) to describe the relationship between an output variable y and three input variables $x_{0,1}$, $x_{0,2}$, $x_{0,3}$, using the attached data set of 427 data points. All data have been normalized such that the variables take values in $[0, 1]$. Consider the following ANN structure, where Layer 0 is the input layer (which is fictitious and each node has the same input and output values), Layer 1 is the hidden layer in which each grey node involves an activation function, and Layer 2 is the output layer which does not have an activation function. The activation function selected is the rectified linear unit (ReLU). The ANN training process is to select appropriate values of the biases (i.e., "$b_{l,j}$") and weights (i.e.,"$w_{l,i,j}$") such that the sum of the squared error is minimized.

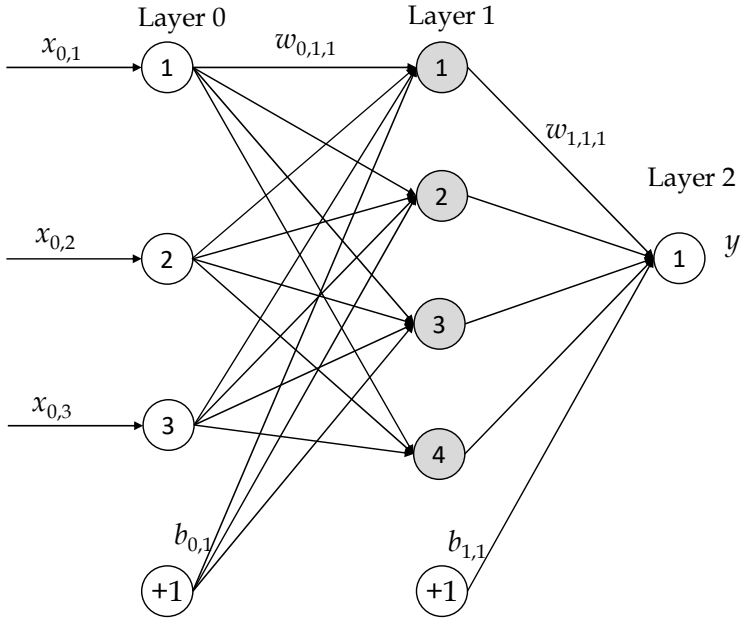

Figure 1: Structure of the ANN to be trained

Please answer the following questions (either in Chinese or in English):

(a) (5 marks) Someone formulated the following mathematical programming problem for training the ANN. Please identify mistakes and incompleteness of the formulation:

$$\min_{w_{lij}, b_{ij}} \quad \sum_{k=1}^{n} (y_k - w_{111} f(w_{011} x_{01,k} + b_{01}) + w_{112} f(w_{012} x_{02,k} + b_{01}) + w_{113} f(w_{013} x_{03,k} + b_{01}) + b_{11})^2$$

where $x_{01,k}$, $x_{02,k}$, $x_{03,k}$, y_k represent the input values and output value of data point k ($k = 1, \cdots, 427$).

(b) (7 marks) Please write the mathematical programming formulation for ANN training without using integer variables. Then transform the formulation into a mixed-integer programming problem that includes a nonlinear objective function and a set of linear constraints.

(c) (8 marks) Let's fix the input weights and the bias of the output lay as

$$(w_{1,1,1}, w_{1,2,1}, w_{1,3,1}, w_{1,4,1}) = (0.5767, 0.1875, -0.1472, 0.1251), \quad b_{1,1} = -0.0895,$$

then the optimization problem in part (b) becomes a mixed-integer quadratic programming (MIQP) problem. Solve the MIQP problem using Gurobi or COPT on Pyomo (which is a Python based modelling platform). See the attached Pyomo guide for more details. Set a solution time limit for the solver (which should be at least 100 seconds). Tell whether a global solution is obtained when the solution procedure terminates, and if not, what is the incumbent solution, the objective value at the incumbent solution, and the lower bound, respectively. Please provide a file that includes the Pyomo/Python codes.

(d) (5 marks) Solve the continuous relaxation of the MIQP problem in part (c). This is a quadratic programming problem (QP) which can be solved by Gurobi or COPT. Compare the solution with the solution of the MIQP obtained in part (c), and discuss the results. Please also provide the relevant codes.

(e) (5 marks) Use PyTorch or any machine learning package to train the ANN using the provided data and the ANN structure in Figure 1. Please provide the command/codes used to train the ANN. Compare the trained ANN with the one obtained in part (c) and discuss the results.