

ORML assignment 1 [34 points]

Formulating a fully connected neural network as an MILP

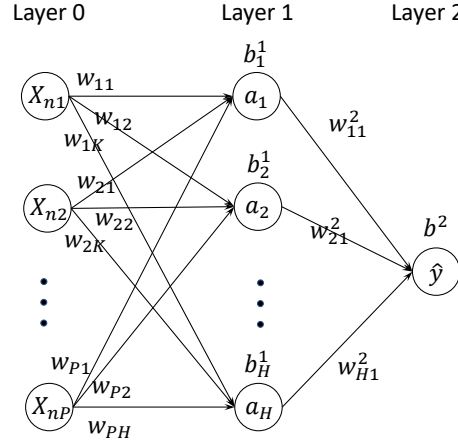


Figure 1: Single-layer neural network along with notation.

Consider a fully connected neural network with 1 hidden layer and H hidden nodes as depicted in Figure 1. The network is used for a binary classification problem, i.e. the output classification $\hat{y} \in \{0, 1\}$. The activation function for each node in the hidden layer is the ReLU function, and the activation function for the output layer is the sigmoid function. The objective of the neural network is to minimize cross entropy.

In this assignment you will optimize the network's hyperparameters for a breast cancer classification dataset (exercise 1), write this feedforward neural network (with some modifications) as an MILP (exercise 2), and compare the performance of the neural network implementation and the MILP (exercise 3). You may use the following notation throughout the assignment:

- the matrix $X \in \mathbb{R}^{N \times P}$ denotes the input data, where element X_{np} corresponds to sample n , feature p ;
- $y \in \{0, 1\}^N$ denotes the vector of true binary classes;
- w_{ij}^k denotes the weight corresponding to the connection from the i^{th} node in layer $k - 1$ to the j^{th} node in layer k , $k \in \{1, 2\}$;
- b_j^k denotes the bias that is added to the linear combination of weights and activations of the previous layer, corresponding to node j in layer k , $k \in \{1, 2\}$;
- $z_{jn} = w_{\cdot j}^1 X_n + b_j^1$;
- a_{jn} is the activation of the j^{th} node in layer 1 for sample n , i.e., $a_{jn} = \text{ReLU}(z_{jn})$;
- $\hat{y} \in \{0, 1\}^N$ denotes the vector of predicted binary classes, i.e. $\hat{y} = \sigma(w_{\cdot 1}^2 a_{\cdot n} + b^2)$

The dataset you'll be using contains information from breast cancer patients [Street et al., 1993], obtained from the UCI repository¹. Based on images of a breast mass, metrics such as tumor radius and compactness have been derived. An overview of all the features can be found in featureNames.csv. There are two types of tumors: malignant (cancerous tumors that spread) and benign (non-cancerous tumors that do not spread). In this assignment you will develop machine learning models that predict whether a tumor is malignant or benign (stored in Y.csv), based on the image-derived metrics (stored in X.csv). Instead of using the data files provided, you can choose to load the data via sklearn².

Exercise 1

[10 points] Search for the best set of hyperparameters, including:

- number of hidden layers;
- number of nodes in the hidden layer;
- batch size;
- regularization;
- optimization algorithm;
- learning rate;
- initialization.

Describe your strategy, which (combinations of) hyperparameter choices you have tried, which evaluation criteria you used for your analysis, and how you came to a final conclusion. Include any graphs or tables that you find useful for your explanation. You may use at most 70 configurations.

Exercise 2

For this exercise two things about the network will be changed. First, the classification is not computed with the sigmoid function, but with the step function:

$$\hat{y} = \begin{cases} 1 & \text{if } w_{j1}^2 a + b_1^2 \geq 0 \\ 0 & \text{if } w_{j1}^2 a + b_1^2 < 0. \end{cases}$$

Second, instead of minimizing cross entropy, we will minimize the number of incorrectly classified samples:

$$\min \sum_{n=1}^N |y - \hat{y}|.$$

1. [4 points] Show that without loss of generality, one can set $w_{j1}^2 = \{-1, 1\}$ for all $j = \{1, \dots, H\}$.
2. [8 points] Write the above described fully connected neural network as an MILP. Use that $w_{j1}^2 \in \{-1, 1\}$.

¹<https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>

²https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html

Exercise 3

[12 points] Train both the neural network and the MILP, where for both you can use the same number of nodes in the hidden layer as you have chosen in exercise 1. Compare the two approaches. What are the (dis)advantages of the neural network, and what are the (dis)advantages of the MILP?

Deliverables

Submit a zip file through canvas containing:

1. the original dataset;
2. a python script called "network.py" of your neural network, along with the code that you used to choose the hyperparameters (this may be in the same script or in a separate script called "hyperparameters.py");
3. a script (python or matlab) of your MILP formulation, called either "MILP.py" or "MILP.m";
4. a report that answers each of the exercises and motivates your choices.

References

W Nick Street, William H Wolberg, and Olvi L Mangasarian. Nuclear feature extraction for breast tumor diagnosis. In Biomedical image processing and biomedical visualization, volume 1905, pages 861–870. SPIE, 1993.