Ted Rogers School of Information Technology Management

ITM 740: Artificial Intelligence in Business

Final Course Project

Deadline: Sunday April 12, 2020

Heart Disease Classification Using Neural Networks

The problem is related to healthcare applications of AI algorithms. The **classification goal** is to predict whether the patient presents heart disease or not. The target class is the last attribute (*Heart disease*) and has two values (0 = no, 1 = yes).

You should split the data into 70% as the training set and 30% as the testing set.

Attribute Information:

1. age: The person's age in years
2. sex: The person's sex (1 = male, 0 = female)
3. cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
4. trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)
5. chol: The person's cholesterol measurement in mg/dl
6. fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)
7. restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
8. thalach: The person's maximum heart rate achieved
9. exang: Exercise induced angina (1 = yes; 0 = no)
10. oldpeak: ST depression induced by exercise relative to rest ('ST' relates to positions on the ECG plot. See more here)
11. slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
12. ca: The number of major vessels (0-3)
13. thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
14. target: Heart disease (0 = no, 1 = yes)

Steps

The project involves the following steps:

1. **Data exploration**: try to know data and represents statistics for the important features among the features and the target attribute. You can use this tutorial to do data exploration in Python: https://www.kaggle.com/pmarcelino/comprehensive-data-exploration-with-python
2. **Use Neural Network and Deep Learning to build a classifier.** Use Neural Network algorithms provided by Keras/TensorFlow to train a model based on training examples. You should try different architectures (e.g., layers, nodes) to achieve the best results.

3. **Test the learned model** on the test set and report the testing results in terms of different parameters such as accuracy, AUC, output error.
4. **Parameter sensitivity analysis:** report results in terms of accuracy when tuning the parameters. For example, presents your observation when you change different parameters such as:
    1. Change number of epochs
    2. Change number of neurons
    3. Use different activation functions
    4. Different optimizer functions

What to hand in:

1- Your programs (**.ipynb files**) for this project.
2- A **report** that contains:
    a) Data exploration to show how the input data look like.
    b) Describe your Neural Network Architecture (in Keras/Tensorflow) to learn the classifier, and the testing results of the learned models (with different settings) on the test data.
    c) Parameter sensitivity analysis that represents the results of different evaluations based on different values of the input parameters of the model. The analysis can be done based on different performance measures such as accuracy, AUC and running time.
    d) Any discussion and conclusion that you find during the experiments.
3- Overview Slides
    a) The slides should present an overview of the project. You can have up to 10 slides.

How to hand in:

1) **Overview Slides**
2) **Project Report**
3) **.ipynb file program**

You must upload your slides (as a PDF file), TensorFlow program (as a text file) and your report (as a PDF file) to D2L by the deadline.

Marking Scheme (25 points)

Your project mark will consist of the following components:

- Your slides (4 points)
- Clearness and organization of your report (4 points)
- Soundness and correctness of your solution (as described in your report and implemented in your programs if any) (7 points).
- Experimental results and parameter sensitivity analysis (10 points)