

SECOND ASSIGNMENT

Developing a Neural Network for solving a Regression problem

Dr Pamela Bezerra

21 of March 2022

PROJECT GOAL: Write a python program to create and test a neural network for solving a regression problem. In your development process, you need to test multiple hyperparameter to obtain the most appropriate network for your problem. The hyperparameters you'll have to analyse are listed below:

1. Learning Rate
2. Batch Size
3. Number of epochs
4. Activation functions
5. Optimizers
6. Loss function
7. Number of Hidden Layers
8. Number of nodes in the hidden layers

OBS_1: when testing one hyperparameter, you need to fix all the others (for example, while testing multiple activation functions, such as ReLu or Sigmoid, all the other parameters are kept the same) to properly analyse the impact of each hyperparameter in the network.

Please use cross-validation to find your best network. To evaluate the impact of different hyperparameter, you can consider the metrics studied in our lectures (precision, recall, accuracy, etc). You can also plot the loss function graph to observe how the loss of the network reduces or increases based on the values of a chosen hyperparameter.

The group is free to use any of the strategies studied in the Lectures to explore different values for hyperparameters (such as Learning Rate Schedule, Random Search, etc). **Please avoid using deep networks.** This problem can be solved with simpler architectures. However, if you observe that your network is diverging or overfitting, feel free to use any of the methods for training deep learning method discussed in previous lectures (such as regularization, weight initialization, etc).

Any python library (scikit-learn, keras, pytorch, tensorflow, matplotlib, seaborn, etc) and IDE (Jupyter Notebook, Google Collab, Spyder, etc) can be used in this project, however the essential steps your code should include are:

- Load the dataset
- Split and clean the dataset
- Load and train the neural network with different hyperparameters
- The many strategies used for testing multiple hyperparameters
- Plot the results for the best network obtained

Please include comments describing each of these key steps in your code. You are free to include any extra steps (such as data visualization) that you like (optional – this won't result in extra points).

DATASET: The dataset used in this project is the House Sales in King County, USA dataset - <https://www.kaggle.com/harlfoxem/housesalesprediction>

This database contains data related to different houses for sale in King County, USA. The price of each house depends on a number of factors, such as the number of bedrooms and the size of the house. The aim of the project is to predict the house price based on these features. A detailed description of each column on the csv file can be found here - <https://www.kaggle.com/datasets/harlfoxem/housesalesprediction/discussion/207885>

REQUIREMENTS: This assignment should be completed in pairs and consists of two deliverables:

- The python code (one or more files in any preferable IDE, such as Jupyter Notebook or Google Collab) (40%)
- A short report of maximum of three pages (not considering cover page and references, in case you want to include) (60%). This report should contain:
 - a. An introduction Section explaining the development process (20%):
 - i. The libraries used
 - ii. The strategies used for testing different values of hyperparameters.
 - iii. The values of the hyperparameters tested.
 - iv. The training and testing process (data split ratio, how the impact of each hyperparameters was evaluated, etc).
 - b. An evaluation section describing your results (60%):
 - i. A brief description of the networks generated and the impact of each hyperparameter, if overfitting or underfitting was observed during the training process, etc.
 - ii. The accuracy results for the best network (or networks, if more than one showed similar results)
 - iii. Your conclusions on the results (e.g., the reasons for this network perform better than all the other trained) based on your AI knowledge.
 - c. Final conclusions (20%)
 - i. The challenges of the project
 - ii. The project task allocation (what each member did)

OBS_2: PLEASE INCLUDE THE NAME AND ID OF BOTH MEMBERS IN THE REPORT. BOTH MEMBERS NEED TO SUBMIT ON CANVAS A ZIP FOLDER CONTAINING THE CODE AND REPORT.

DEADLINE: WEDNESDAY, 6TH OF APRIL 2022 AT 17:00

GUIDANCE: Please consult the following materials for further guidance on implementing Neural Networks in python:

1. Python Data Science Handbook - <https://jakevdp.github.io/PythonDataScienceHandbook/>
2. Hands-on Machine Learning with scikit-learn, keras and tensorflow (book) - <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/> - Especially Chapters 10 and 11
3. Pytorch tutorials on Neural Networks - https://pytorch.org/tutorials/beginner/blitz/neural_networks_tutorial.html
4. TensorFlow tutorial on Neural Networks for beginners - <https://www.tensorflow.org/tutorials/quickstart/beginner>

5. Neural Networks and Deep Learning (book) -
<http://neuralnetworksanddeeplearning.com/chap3.html> - especially Chapter 3

Good luck to everyone,

Best Regards,

Pamela Bezerra