# Neural Network; Regression Task; Example in Python; AMS 580

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Please include (1) Python file; (2) Output from Python with answers to all the questions asked; (3) Comparison of the results to those using R; (4) Recommended websites for Neural Network using Python.

#### Neural Network with the Boston Housing price Data – Regression Task

The Boston housing price data we will use for our homework is taken from the Kaggle competition site (<https://www.kaggle.com/vikrishnan/boston-house-prices>). **We will treat this dataset as our entire data.**

Each record in the database describes a Boston suburb or town. The data was drawn from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. Our data has 506 observations and 14 variables:

Attribute Information (in order):

**CRIM:** per capita crime rate by town

**ZN:** proportion of residential land zoned for lots over 25,000 sq.ft.

**INDUS:** proportion of non-retail business acres per town

**CHAS:** Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

**NOX:** nitric oxides concentration (parts per 10 million)

**RM:** average number of rooms per dwelling

**AGE:** proportion of owner-occupied units built before 1940

**DIS:** weighted distances to five Boston employment centers

**RAD:** index of accessibility to radial highways

**TAX:** full-value property-tax rate per 10,000 dollars

**PTRATIO:** pupil-teacher ratio by town

**Black:** The result of the equation B=1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

**LSTAT:** %lower status of the population

**MEDV:** Median value of owner-occupied homes in 1000 dollars

First, one must clean the data and decide which variables to exclude from our analysis. My recommendation is that we exclude *RAD* in the ensuing analysis. Now after the data cleaning step, your task is to split the data randomly into training (75%) and testing (25%), first build the best neural network models to predict median house value using the training data, and then use these models to predict the house price in the testing data. Please note that we usually use the *neuralnet* package in R to build the various neural network models when the activation function is the same across layers, while we usually use the *keras* package when the activation functions are not the same.

**Note:** For this data set, we shall NOT perform data standardization (normalization) – just to make it easier for you. In the future, for your real-world application, please do consider normalization for all machine learning methods, including the neural network.

Please review the following websites for related methods and procedures in R:

1. **neuralnet**:

<https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf>

<https://datascienceplus.com/fitting-neural-network-in-r/>

<https://rpubs.com/shyambv/linear_neural_network>

2. **multiple linear**:

<http://www.sthda.com/english/articles/40-regression-analysis/165-linear-regression-essentials-in-r/>

1. For the entire dataset, please perform the data cleaning as instructed before; namely, exclude the variables *rad*. Then please use the random seed 123 to divide the cleaned data into 75% training and 25% testing.

1. Please first build the predictive model to predict the house price using the training data and the NN model with (i) no hidden layer, (ii) the default loss function of “sse”, and (iii) the default activation function of “identity”. Please plot the model obtained using the training data. Please compute the Test MSE using the testing data.
2. Please first build the predictive model to predict the house price using the training data and the NN model with (i) one hidden layer with 3 neurons, (ii) the default loss function of “sse”, and (iii) the default activation function of “identity”. Please plot the model obtained using the training data. Please compute the Test MSE using the testing data.
3. Now we shall build the predictive model to predict the house price using the training data and the Multiple Linear Regression model. Please report the fitted linear regression model obtained using the training data – and compare to the NN model in part 2 and 3. Please compute the Test MSE using the testing data.
4. Please first build the predictive model to predict the house price using the training data and the NN model with (i) one hidden layer with 3 neurons, (ii) the default loss function of “sse”, and (iii) the output layer with the default activation function of 'identity', but the hidden layer with the activation function of 'relu'. Please plot the model obtained using the training data. Please compute the Test MSE using the testing data. Please add the predicted class label to the testing dataset.

