Homework 3 – Feature Selection and Artificial Neural Network

Problem 1

The dataset "`heating_load.csv`" presents some energy consumption data for a series of buildings. All the buildings have the same overall volume (771.75 m^3) but different shapes (e.g., different surface, height, etc.). The dataset contains the following information:

1) **Input features:**
    a. Column 1: Total surface area (in m^2)
    b. Column 2: Total area of the walls (in m^2)
    c. Column 3: Total area of the roof (in m^2)
2) **Output label:**
    a. Column 4: Measured heating load (in BTU, "British Thermal Unit")

Train a multivariate linear regression model aiming to predict the buildings' heating load in terms of the three input features characterizing the geometry of the buildings. Try engineering new features (by transforming and/or combining the initial features). Use the LASSO regularization method to select the optimal features. Compare the validation set RMSE offered by linear regression and LASSO.

Problem 2

Using the same dataset, train an artificial neural network (ANN) regression model aiming to predict the buildings' heating load in terms of the three input features characterizing the geometry of the buildings. Determine the optimal architecture for the ANN model (e.g., number of hidden neurons). Compare the validation set RMSE offered by ANN with those yielded by linear regression and LASSO and explain in the report why there exists such a difference.