

CSE/ISE 337: Scripting Languages

Stony Brook University

Programming Assignment #5

Spring 2024

Assignment Due: Sunday, May 5th, 2024, by 11:59 PM (EST)

Problem 1 [50] : (Anusha, Maneesh and Preethika)

This exercise will guide you through the process of creating a **linear regression model** for data about a fish market. You can use the code from the class lectures on Pandas and Linear Regression, especially to clean the data and identify the correlations.

Read the data

1. You will do this problem in a Jupyter Notebook and submit a Jupyter Notebook file, Problem1.ipynb
2. You have been given the data "Fish.csv"
3. You have been given the file "info.txt"
4. You have been given the file "Problem1.ipynb"

Clean the data

1. Open the info.txt file and read the info for each column.
2. Rename the Length1 column to VerticalLength, Length2 to DiagonalLength, and Length3 to CrossLength,

Identify correlations

1. Use the corr() method to look at the correlation data for the Weight column.
2. Create a heatmap for the correlation data

Create a regression model

1. Use the correlation data to choose three independent variables. The first two should be the Height and Width columns. The third should be one of the length columns
2. Filter the DataFrame so it only contains the values *for* the Bream species and assigns the resulting DataFrame to a variable named Bream.
3. Use the train_test_split() method to split the training and testing data for the **Bream** species data.
4. Create the model and fit the training data.

5. Score the testing data.
6. Make the predictions and store the results in a DataFrame
7. Join the column for the predicted data with the columns for the actual data and assign the resulting DataFrame to a variable named `final`.

Plot the residuals

1. Calculate the residuals and store them in a new column in the DataFrame.
2. Use the `displot()` method to create a displacement `plot` that plots the residuals..

Problem 2 [50]: (Dheeraj and Pritish)

This exercise will guide you through the process of creating a **Neural Network using Pytorch** for data about a car market. You can use much of the code from the class lecture on PyTorch. For this problem, you need to understand and study the model covered in the class.

Read the data

1. You will do this problem in a Jupyter Notebook and submit a Jupyter Notebook file, Problem2.ipynb
2. You have been given the data "cars.csv"
3. You have been given the file "Problem2.ipynb"
4. If necessary, run the Kernel→Restart Kernel and Clear All Outputs command.

Build a Model

1. Create a NN model
 - a. There should be three hidden layers
 - b. The outer layer (h1) should have 9 neurons
 - c. The middle layer (h2) should have 8 neurons
 - d. The last layer (h3) should have 7 neurons
 - e. They should be fully connected
 - f. You can select any activation function
 - g. The input feature layer size depends on the number of features you will select from the data after co-relation analysis
 - h. The output layer depends on the number of classes in your data

Clean and prepare the data (training and testing)

1. Clean the data. You will be building a NN to classify a car for carbody (Convertible, hatchback, sedan)
2. Determine the co-relation of features with respect to the carbody
3. Select the top five features to represent a car object/data
4. Train and split the data for training and testing
5. Select a loss function
6. Select an optimizer
7. Select epochs (min 100)
8. Training and test data
9. Test the model and print the loss performance

Plot the Traing Graph

1. Plot a graph between loss and epoch

Save the Model

1. Save the trained model.