**Neural Network Deep Learning**

Assignment – 4

Name: Kishor Kumar Andekar

Student ID: 700744713

Github Link : <https://github.com/kishorreyansh/Neural-Network-Deep-Learning/tree/main/Assignment-4>

1. Data Manipulation

a. Read the provided CSV file ‘data.csv’.

b. <https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>

c. Show the basic statistical description about the data.

d. Check if the data has null values.

i. Replace the null values with the mean

e. Select at least two columns and aggregate the data using: min, max, count, mean.

f. Filter the dataframe to select the rows with calories values between 500 and 1000.

g. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.

h. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.

i. Delete the “Maxpulse” column from the main df dataframe

j. Convert the datatype of Calories column to int datatype.

k. Using pandas create a scatter plot for the two columns (Duration and Calories).

In the below code snippet, we are doing data manipulation using Pandas:

Reading the provided CSV file ‘data.csv' using the read\_csv function, assigning a variable df\_dataframe to it, and showing the statistical description of the dataframe using the description() function. Replacing the null values in all the columns with the mean. Selecting two columns (duration and pulse) and aggregating the data using the.agg() function and using separate variables to filter the dataframe to select the rows with calorie values between 500 and 1000 and to select the rows with calorie values > 500 and pulse <100. Next, create a new “df\_modified” dataframe that contains all the columns from the dataframe except for “Maxpulse” and delete the “Maxpulse” column from the main df\_dataframe dataframe. Converting the datatype of the Calories column to an int datatype using the.astype() function and finally using Pandas to create a scatter plot for the two columns (Duration and Calories).

A screenshot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

Output:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screen shot of a graph

Description automatically generated

1. Linear Regression
2. Import the given “Salary\_Data.csv”
3. Split the data in train\_test partitions, such that 1/3 of the data is reserved as test subset.
4. Train and predict the model.
5. Calculate the mean\_squared error
6. Visualize both train and test data using scatter plot

In the below code snippet, we are training data using linear regression:

Importing the given “Salary\_Data.csv” into a variable called df\_dataframe. Splitting the data using the train\_test\_split() function, such that 1/3 of the data is reserved as a test subset. Train the model using the LinearRegression() function, predict the values using the predict() function, calculate the mean squared error of the predicted set, and use Pandas to visualize both the train and test data using a scatter plot.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

Output:

A screenshot of a computer program

Description automatically generated

A screen shot of a graph

Description automatically generated