Linear Regression Assignment

We have abalone dataset from UCI machine learning repository.

Perform exploratory data analysis on this dataset

1. Import and check the shape of data

2.remove the attribute ‘unamed: 0 ’

3. plot the correlation matrix of each attribute using heatmap of seaborn

4. Fit a linear line of age with respect to length, weight, diameter and other attributes and plot the di stribution of each attribute with respect to age.

5. Create a boxplot of age and sex

6. Create a boxplot of whole dataset

7. Dummy Sex attribute

8. Try different splits for train and test data

9. Create multiple models with different features for predicting the age attribute and compare each model’s rmse.

For ex: take independent variables as height, Wholeweight, Diameter and dependent variable as age and create a linear model.

10. Check the evaluation metrics for each model (rmse)

11.Analyse the result.

The description of the data can be found at:

http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.names

Solutions to this question paper will be uploaded on lms after 2 days