School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET301
Year- 2022
Date- 19-09-2022

Type- Core
Course Name-AIML
Semester- Odd
Batch- V Sem

Lab Assignment No. 5.2.2

Exp. No.	Name	CO-1	CO-2	CO-3
5.2.2	SVR (Regression Problem)	✓	✓	--

Objective: To implement Support vector Regressor (*SVRs*) (using Scikit-learn)

Download the dataset from
https://drive.google.com/file/d/1krxrtHmAwwNmreefnhDp6Tn0cmafQG7g/view?usp=sharing (5)

Problem Statement:

A Chinese automobile company Geely Auto aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts.

They have contracted an automobile consulting company to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market. The company wants to know:

- Which variables are significant in predicting the price of a car
- How well those variables describe the price of a car

Based on various market surveys, the consulting firm has gathered a large dataset of different types of cars across the American market.

You are required to model the price of cars with the available independent variables. It will be used by the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

1. Read the dataset. (5)
2. Check the number of columns in the dataset. (3)
3. Display 10 rows of the dataset. (2)

4. Check type of each attribute. (2)
5. Check the presence of missing values and delete the instances (rows) where more than 60% data are missing. (3)
6. Check the presence of Categorical columns. Handle it if present. i.e., Transform categorical features into numerical features. (Hint: Use eitherone hot encoding, label encoding or any other suitable pre-processing technique). (10)
7. **Split** the dataset into **80% for training** and rest **20% for testing** (10) (sklearn.model_selection.train_test_split function)
8. Scale the numerical columns value using MinMaxScaler. (5)
9. **Run SVR Regressor** using built-in function on the training set with default parameters (sklearn.svm.SVR)(10)
10. Check the performance of the trained model on test set. Find the value of R^2 score. (5)
11. Play with the different parameters of SVR from its documentation (https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html). (10)
12. Compare the results of SVR regressor with previously used MLR. (20)

Suggested Platform: Python: Jupyter Notebook/Azure Notebook/Google Colab.