REPORT

Princy Gautam(B20BB051)

METHOD

- Import necessary libraries for performing calculations and reading the csv file.
- Mount drive on colab and read the csv data file using pandas library.
- Check the shape and head of the dataset.
- Chose the three columns ‘Age’, ‘Income’ and ‘Securities’ as features and ‘CreditCard’ as the target and separate them in variables
- Check for missing values in the dataset.
- Plot the 3D scatter plot giving features as the input.
- Split the dataset in the ratio of 80:20
- Train the model with LinearSVC and set the regularization parameter C as a set of number. Fit the model on training data.
- Made predictions on test data and printed the score for each value of C.
- Print confusion matrix and classification report.
- Apply a cross validation method to find the most appropriate value of C. Store it in a variable named as grid. Fit the model on training data and print the best value of C.

RESULT and OBSERVATIONS

- Shape of dataset is (5000,7)
- There are no missing values in the data.
- The 3D scatter plot shows the relationship between the three variables, with one variable each axis.
- The dataset is split into (4000,4) and (1000,4)
- The score on test data for each value of C is 1.0
- Using GridSearchCV we get the best value of C as 0.001
- We get an accuracy of 1.0 after applying regularization hyperparameter. Thus, it helps to increase the accuracy of the model and controls the learning process.

Colab file link:
https://colab.research.google.com/drive/11dZzdARKqv92sTZG0UGHibPfxYRA4j0z#scrollTo=es-uWy_uIkYQ