INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ASSIGNMENT 2
(MACHINE LEARNING)

Gandham Heamanth Rao 20CS10027

Talabattula Sai Sahan 20CS30055

SUPERVISED LEARNING

Question:

1) Normalise the data using Standard Scalar Normalisation. Randomly divide the Dataset into 80% for training and 20% for testing. Encode categorical variables using appropriate encoding method (in-built function not allowed for normalization, sampling and encoding).

2) Implement the binary SVM classifier using the following kernels: Linear, Quadratic, Radial Basis function. Report the accuracy for each. (in-built function allowed).

3) Build an MLP classifier (in-built function allowed). for the given dataset. Use stochastic gradient descent optimiser. Keep learning rate as 0.001 and batch size of 32. Vary the number of hidden layers and number of nodes in each hidden layer as follows and report the accuracy of each:

a. 1 hidden layer with 16 nodes

b. 2 hidden layers with 256 and 16 nodes respectively.

4) Using the best accuracy model from part 3, vary the learning rate as 0.1, 0.01, 0.001, 0.0001 and 0.00001. Plot the learning rate vs accuracy graph.

5) Use backward elimination method on the best model found in part 3 to select the best set of features. Print the features.

6) Apply ensemble learning (max voting technique) using SVM with quadratic, SVM with radial basis function and the best accuracy model from part 3. Report the accuracy.

SOLUTION:

Because of using random state in splitting i have pasted one of the result during implementation.

Every accuracy is maintained in fractions.

The results obtained are given below:

###########normalization of data started##########
###############Linear SVM classifier#############
SVM accuracy for linear kernel: 0.9666666666666667

###############Quadratic SVM classifier##########
SVM accuracy for quadratic kernel: 0.8333333333333334

##################Radial SVM classifier##########

SVM accuracy for radial kernel: 0.9666666666666667

###############MLP classifier####################
The mlp accuracy for batch size 32, with 1 hidden layer and 16 nodes: 1.0
The mlp accuracy for batch size 32, with 2 hidden layers - 256 and 16 respectively:
0.9666666666666667
#############learning rate vs Accuracy##############
learning rate   -   accuracy
0.1   -   0.9333333333333333
0.01   -   0.9
0.001   -   0.9
0.0001   -   0.9
1e-05   -   0.9

############################The Best Features#########################
The Best Features Are: Index(['petal length in cm'], dtype='object')
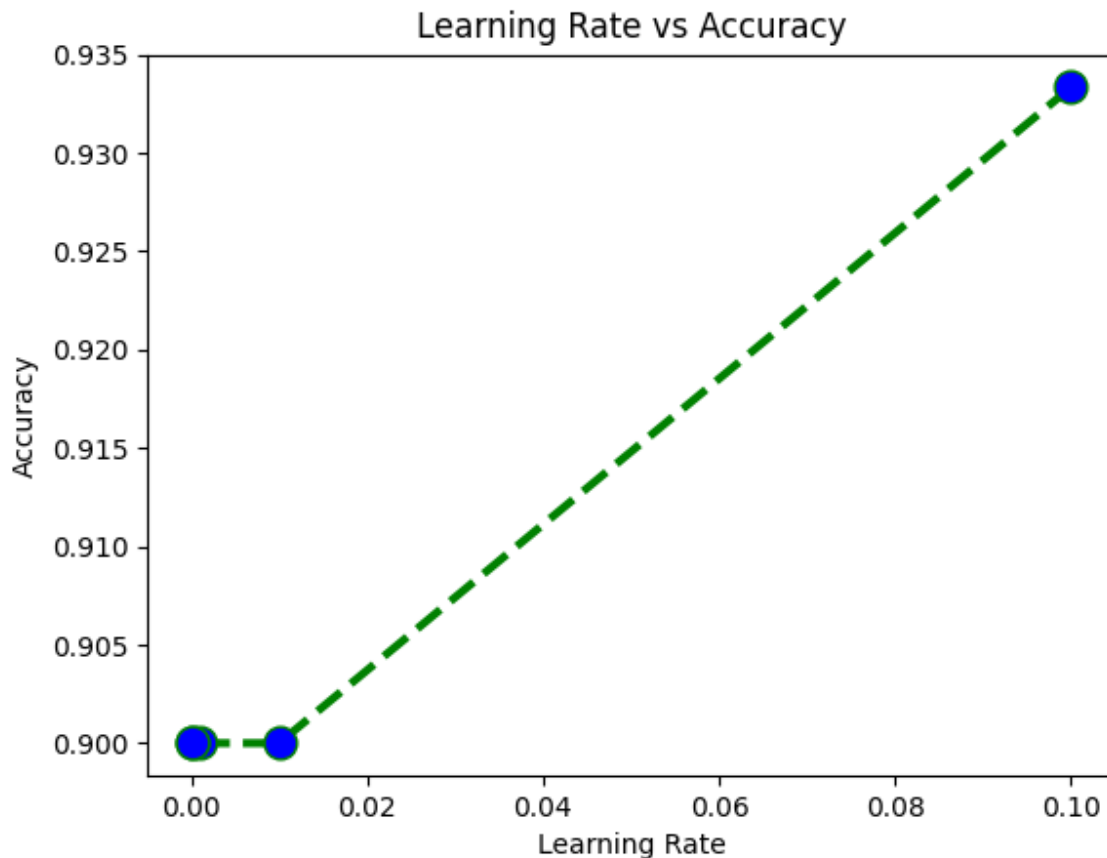Accuracy with above Features: 0.9666666666666667

##########################Maximum Vote
Classifier##############################
Accuracy for Max vote classifier - 0.9666666666666667

Below I have plotted the learning rate vs Accuracy.

<u>LEARNING RATE vs ACCURACY</u>

Learning Rate vs Accuracy

ALGORITHM DONE:

1. Normalized using normalize function which I wrote a function for that.
2. I have split the data into 80-20 ratio.
3. I have done SVM, MLP using inbuilt and plotted the function.
4. I have selected the best features using back elimination by eliminating each column recursively .
5. Maximum voting is done using algorithm which has max_classifier and done by highest voting rule.

OBJECTIVES DONE:

1. Has normalized the given data.
2. Have done SVM, MLP and applied the learning rate vs accuracy for the best in 3rd question.
3. Has done backward elimination and printed the best features.
4. Has done the max vote and printed the accuracy.