

CSE 572

Data Mining

Instructor- Prof. Arunabha Sen

Assignment 5

Total marks: 20

Submission Deadline : 04.23.2019 11:59 pm AZ time.

- For submission, you should submit codes and a PDF report containing the results in a zipped file (only one submission per group). The PDF file should contain names of all the members. The zipped file name should be in the following format:

GroupName_GroupID.zip [eg.- DM_12.zip]

- Refer to the ‘group formation sign-up’ sheet in the blackboard for group ID and GroupName (Group name should be the First name of Member 1. Group ID can be obtained from the first column).
 - For coding you can use both Matlab and Python. For Matlab and Python codes, include .m and .py files in the zipped folder respectively.
-

In this Assignment, you are required to build models for (a) regression and (b) classification

You are allowed to use Naïve-Bayes Classifier and SVM libraries (eg.- learn.naive_bayes and sklearn.svm packages when using Python).

Task 1)

[5 marks]

Naïve-Bayes Classifier

In this classification problem, you are required to train a Naïve-Bayes classifier that predicts whether a person is male (represented as 0) or female (represented as 1), given three features: height (in centimeters), age and weight (in kilograms).

Use “PB1_train.csv” and “PB1_test.csv” for this task, where the first three columns represent three features (height, age, weight), and the fourth column represent class label (0/1). Train a Naïve-Bayes classifier NV on “PB1_train.csv” data that learns to map the mentioned features to their corresponding class values.

Report the predicted values (\bar{y}) and accuracy percentage (percentage of matches) of the model NV by testing it on “PB1_test.csv” data.

[1] Accuracy (in percentage) , [2] Predicted Values (\bar{y}) on “PB1_test.csv”.

Task 2)

[5 marks]

SVM Classifier

Use “PB1_train.csv” and “PB1_test.csv” for this task. Train three SVM classifiers SV1, SV2 and SV3 on “PB1_train.csv” data that learns to map the mentioned features to their corresponding class values.

Report the predicted values (\bar{y}) and accuracy percentage (percentage of matches) for each model by testing it on “PB1_test.csv” data.

SV1 configuration: kernel type = 'linear'.

SV2 configuration: kernel type = 'polynomial', Degree of the polynomial kernel function = 5.

SV3 configuration: kernel type = ‘radial basis function’ or 'rbf'.

[1] Accuracies (in percentage), [2] Predicted Values (\bar{y}) on “PB1_test.csv” for each model.

Task 3)

[5 marks]

Naïve-Bayes Classifier

Use “PB2_train.csv” and “PB2_test.csv” for this task, where the first three columns represent three features (height, age, weight), and the fourth column represent class label (0/1). Train a Naïve-Bayes classifier NV on “PB2_train.csv” data that learns to map the mentioned features to their corresponding class values.

Report the predicted values (\bar{y}) and accuracy percentage (percentage of matches) of the model NV by testing it on “PB2_test.csv” data.

[1] Accuracy (in percentage), [2] Predicted Values (\bar{y}) on “PB2_test.csv”.

Task 4)

[5 marks]

SVM Classifier

Use “PB2_train.csv” and “PB2_test.csv” for this task. Train three SVM classifiers SV1, SV2 and SV3 on “PB2_train.csv” data that learns to map the mentioned features to their corresponding class values.

Report the predicted values (\bar{y}) and accuracy percentage (percentage of matches) for each model by testing it on “PB2_test.csv” data.

SV1 configuration: kernel type = 'linear'.

SV2 configuration: kernel type = 'polynomial', Degree of the polynomial kernel function = 7.

SV3 configuration: kernel type = 'radial basis function' or 'rbf'.

[1] Accuracies (in percentage), [2] Predicted Values (\bar{y}) on “PB2_test.csv” for each model.