Problem 5 (10 pts)

To answer a numerical value, enter only the numerical value in the answer box and units are not required. If a problem statement says “up to N decimal places”, answer by rounding off the N+1 decimal places.

Spam (class 1) and non-spam (class 0) comments are recorded in dm_mid5.csv. Use this to create a Naive Bayes classifier under the following conditions:

To generate training and test data, use sklearn.model_selection.train_test_split() function with arguments test_size=0.25, random_state=20. The Naive Bayes classifier should be created with the training data in the Bag of Words representation, with smoothing parameter α=1 (Laplace smoothing).

Answer the number of training data, the number of test data, and the identification rate of the test data (ratio from 0 to 1. up to the third decimal place). If necessary, dm_mid5_sample.ipynb may be used.

the number of training data

the number of test data

the identification rate of the test data

upload ipynb file:

参照... ファイルが選択されていません。

upload html file :

参照... ファイルが選択されていません。