

## 問5 (10点)

数値を答える場合は、数値のみを解答欄に記入すること。単位などは不要。また、小数第N位まで、と書いてある場合は、小数第N+1位を四捨五入して答えること。

dm\_mid5.csvにはスパムコメント(class 1)と非スパムコメント(class 0)が記録されている。これを用いて、以下の条件でナイーブベイズ分類器を作成せよ。

sklearn.model\_selection.train\_test\_split() 関数に、test\_size=0.25, random\_state=20を指定して訓練データとテストデータを作成すること。訓練データからbag of words表現を作成し、ナイーブベイズ分類器を作成する。このとき、スムージングパラメータ $\alpha$ は1にすること。

訓練データ数、テストデータ数、テストデータの識別率 (0から1までの割合。小数第3位まで) を答えよ。必要に応じて、dm\_mid5\_sample.ipynbを用いて良い。

訓練データ数

テストデータ数

識別率

ipynbファイルのアップロード:

参照... ファイルが選択されていません。

htmlファイルのアップロード:

参照... ファイルが選択されていません。