

第 7 章

SVM

7.1 目的

SVM を用いた文書分類を行います。

実践演習 7-1

Weka の SMO で例題 7.3 に基づいて ReutersGrain の学習と評価を行え。その際、カーネルの種類やハイパーパラメータをいくつか変化させて、識別率への影響を観察せよ。

余裕があれば、StringToWordVector フィルタ適用時に、(1) stopwordsHandler の値を Rainbow にする、(2) IDFTrandform, TFTransform を True にする、などの変更を行って、変化を観察せよ。

実践演習 7-2

実践演習 7-1 の内容を sklearn の SVC を用いて行え。その際、(1) CountVectorizer で stop_words 引数 (値は'english') を用いる、(2) TfidfVectorizer を用いるなどして、単語ベクトルを変更して変化を観察せよ。