



Feedback and Classification

Please note: even though the homework assignments are optional, you are encouraged to answer them, as they will help you prepare for your final exam. In case of questions reach out to email: usman@ifis.cs.tu-bs.de.

1. Explain what is the idea behind Rocchio's algorithm for pseudo relevance feedback?
2. Consider Table I that shows the terms representations of six documents. Two classes are shown: USA and Germany.
 - a. Use these values with Naïve Bayes to determine the class for documents d7 and d8.

Table I. Terms representation and class assignments

Document		Terms in Document (1: present, 0: not_present)					
Vector	Class	Dallas	München	Heidelberg	Cambridge	Washington	Boston
d1	USA	1	0	0	1	1	0
d2	USA	1	0	0	1	1	1
d3	USA	0	1	1	1	0	1
d4	Germany	0	1	0	0	0	0
d5	Germany	1	1	1	0	0	0
d6	Germany	0	1	1	0	0	0
d7	?	0	1	1	0	0	1
d8	?	1	1	0	1	0	1