

Homework 10

Student Number:

Name:

Problem 1. (30 points) The rationale for the positional independence assumption is that there is no useful information in the fact that a term occurs in position k of a document. Find exceptions. Consider formulaic documents with a fixed document structure.

Problem 2. (40 points) Which of the following documents have identical and different bag of words representations for (i) the Bernoulli model (ii) the multinomial model? If there are differences, describe them.

- 1) He moved from London, Ontario, to London, England.
- 2) He moved from London, England, to London, Ontario.
- 3) He moved from England to London, Ontario.

Problem 3. (30 points) Assume a situation where every document in the test collection has been assigned exactly one class, and that a classifier also assigns exactly one class to each document. This setup is called one-of classification . Show that in one-of classification (i) the total number of false positive decisions equals the total number of false negative decisions and (ii) microaveraged F_1 and accuracy are identical.