Midterm for CMSC 644/498U

April 2, 2018

1 Problem 1

1.1 Part A

Please give the 7-shingles for the following two sentences (treat blank and single quotation mark as single characters). What is their Jaccard similarity?

- S_1 = I drink alcoholic drinks
- S_2 = I don't drink alcoholic drinks

1.2 Part B

What about this pair? Please do the same task as Part A on the following pair of sentences.

- S_1' = I drink milk but I don't drink alcoholic drinks
- S_2' = I don't drink milk but I drink alcoholic drinks

1.3 Part C

What did you learn? What shingle length do we need to distinguish S_1' and S_2'?

2 Problem 2

Suppose we are given two binary random vectors a and b of dimension 16. Each entry is either "0" or "1". Suppose we now decompose each vector into 8 sub vectors of size 2 each; call them $a_1, a_2, \ldots a_8$ and $b_1, b_2, \ldots b_8$.

Example: if $a = (0,0,0,0,1,1,0,0,0,1,0,1,0,0,1,1)$ then $a_1 = (0,0), a_2 = (0,0)$ and $a_3 = (1,1)$ etc.

2.1 Part A

What is the probability that for a specific i, $a_i = b_i$?

2.2 Part B

What is the probability that for at least one value i, $a_i = b_i$?

3 Problem 3

Suppose we are seeing cars of different colors drive down Rt 1. Assume that there is one *dominant* color (in other words strictly more than half the cars have that color). How can we design a *small space* streaming algorithm to identify the dominant color? Assume that the number of possible colors is so large that we cannot possibly keep a count of every possible color. Let us assume that we observe the cars in a 24 hour period and hence we do not know exactly how many cars we will see.

Describe your algorithm in pseudocode, and explain why it works.

4 Problem 4

Fomulate the following problem as a Linear Program (LP).

In class, we mentioned that we can find a maximium weighted matching in a bipartite graph. Please formulate an LP for the following matching problem. Remember to explain briefly the variables and why it works. (You do not need to argue that we can always obtain an integer solution.) Assume e_{ij} has weight w_{ij}.

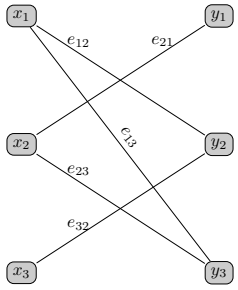

Figure 1: Graph