

2IMW30 - FOUNDATION OF DATA MINING
Assignment 1a Report

Duy Pham (0980384)
Mazen Aly(—)

EXERCISE 2

In this exercise, we are proving that if the Jaccard similarity of two sets is 0, then minhashing always gives a correct estimate of the Jaccard similarity.

If the Jaccard similarity of two sets is 0, then the intersection of the two sets is empty. This means there are no rows which the elements are both "1" in the signature matrix. Hence, when performing minhashing, there cannot be any row with the same values in 2 cells. Thus $P(h(S_i) = h(S_j)) = 0$, which is the same as the Jaccard similarity.