IR (Chapter 19) Classwork Solution

Dell Zhang
Birkbeck, University of London

1. The two hash functions define the following permutations:
   $h_1(0) = 1$, $h_1(1) = 3$, $h_1(2) = 0$, $h_1(3) = 2$, $h_1(4) = 4$;
   $h_2(0) = 1$, $h_2(1) = 4$, $h_2(2) = 2$, $h_2(3) = 0$, $h_2(4) = 3$.

   For any set S define $\min^h(S)$ to be the minimal member of S with respect to h — that is, the member x of S with the minimum value of $h(x)$.

   For $D_1 : \{0, 1, 2\}$,
   $\because h_1(2) < h_1(0) < h_1(1)$, $\therefore \min^{h_1}(\{0, 1, 2\}) = 2$;
   $\because h_2(0) < h_2(2) < h_2(1)$, $\therefore \min^{h_2}(\{0, 1, 2\}) = 0$;
   Therefore its sketch is [2, 0].
   For $D_2 : \{1, 3, 4\}$,
   $\because h_1(3) < h_1(1) < h_1(4)$, $\therefore \min^{h_1}(\{1, 3, 4\}) = 3$;
   $\because h_2(3) < h_2(4) < h_2(1)$, $\therefore \min^{h_2}(\{1, 3, 4\}) = 3$;
   Therefore its sketch is [3, 3].
   For $D_3 : \{0, 2, 3\}$,
   $\because h_1(2) < h_1(0) < h_1(3)$, $\therefore \min^{h_1}(\{0, 2, 3\}) = 2$;
   $\because h_2(3) < h_2(0) < h_2(2)$, $\therefore \min^{h_2}(\{0, 2, 3\}) = 3$;
   Therefore its sketch is [2, 3].

   The pairwise Jaccard coefficients can be estimated as
   $\hat{J}(D_1, D_2) = 0/2 = 0.0$;
   $\hat{J}(D_2, D_3) = 1/2 = 0.5$;
   $\hat{J}(D_3, D_1) = 1/2 = 0.5$.