

MinHash sketches

As discussed in class, for a min-wise independent family H , we can associate a sketch $s(X) = [\min h_1(X), \min h_2(X), \dots, \min h_k(X)]$ with each set X in the given data collection, where h_1, h_2, \dots, h_k are independently chosen at random from H . Consider now any two sets A and B , with their sketches $s(A)$ and $s(B)$. Can you compute a sketch for $A \cup B$ using just $s(A)$ and $s(B)$ in $O(k)$ time? Can you prove that it is equivalent to compute $s(A \cup B)$ from scratch directly from $A \cup B$?

SOLUTION