

NIT Delhi
Assignment-1
Optimization Techniques

Note: Attempt all questions.

1. Suppose we have a universal set U of n elements, and we choose two subsets S and T at random, each with m of the n elements. What is the expected value of the Jaccard similarity of S and T ?
2. Using the data from Fig. 1, add to the signatures of the columns the values of the following

<i>Row</i>	<i>S₁</i>	<i>S₂</i>	<i>S₃</i>	<i>S₄</i>
0	1	0	0	1
1	0	0	1	0
2	0	1	0	1
3	1	0	1	1
4	0	0	1	0

Figure 1:

hash functions:

(a) $h_1(x) = 2x + 4 \bmod 5,$

(b) $h_2(x) = 3x - 1 \bmod 5$

3. Evaluate the S -curve $1 - (1 - s^r)^b$ for $s = 0.1, 0.2, \dots, 0.9$, for $r = 3$ and $b = 10$.
 For each of the (r, b) pairs in Exercise 3.4.1, compute the threshold, that is, the value of s for which the value of $1 - (1 - s^r)^b$ is exactly $1/2$. How does this value compare with the estimate of $(1/b)^{1/r}$.
4. Prove that the cosine distance between any two vectors of 0's and 1's, of the same length, is at most 90 degrees.