Programming and Classification 2021

EXAM

1. Find the centroid of the set

$$\{(0,1,0),(-1,-1,-1),(-1,0,-1),(0,0,1)\}$$

   with Euclidean distance.

2. Construct the characteristic matrix for the family of sets

$$\{\{a,b,c,g\},\{a,b,e,g\},\{a,b,c,f\}\}\ .$$

3. Let S_1 be the set of 3-shingles of the word `Makka` and let S_2 be the set of all 3-shingles of the word `Pakka`. Find $J(S_1,S_2)$ where J is the Jaccard similarity.

4. Consider a set $S = \{$ `xy, y, yc, yx, xx` $\}$ with edit distance. Find the diameter of S.

5. How to conctruct $(5,10,0.75,0.84)$-sensitive functions from $(5,10,0.5,0.6)$-sensitive functions?

6. We know that bags A_1 and A_2 have 10 and 20 elements, respectively. Moreover, we know that there is an element x such that $x \in A_1$ and $x \in A_2$. Provide exact lower and upper bounds on the Jaccard similarity between A_1 and A_2.

7. Two long texts have different characters on 10% positions. We choose 20 positions at random and check if characters placed there are different in both texts. We accept texts as equal if and only if both texts agree on all 20 places. Find the false acceptance rate.

8. When TF.IDF is positive?