

# Assignment 4: Metrics on Features

Lecturer: Dr. Wan-Lei Zhao

Xiamen University

## 1 Problem-1

min-Hash [1] is an efficient way to approximate the *Jaccard* distance between two sets. Please implement min-Hash in C++ to approximate the pair-wise *Jaccard* distances between the provided sets. Please try different numbers of the hashes in your experiment, and then show the variation trend of the approximated *Jaccard* distances in your report.

In the attached file “data/sets.txt”, you are provided with ten sets of English Characters in the range of [‘A’, ‘L’]. Please write code (in Python/C++) to calculate the pairwise Jaccard distance between these sets based on the permutations defined in Fig. 1.

## 2 Problem-2

Given 64-bits binary code, both of which are kept as an unsigned long integer. Please find out the efficient way (in C++) to calculate their Hamming

	A	B	C	D	E	F	G	H	I	J	K	L
$\pi_1$	5	11	12	9	1	6	3	2	7	10	4	8
$\pi_2$	5	11	2	3	6	9	10	7	8	1	4	12
$\pi_3$	6	8	9	7	2	3	12	4	11	1	5	10
$\pi_4$	8	6	4	1	5	7	11	12	9	2	3	10
$\pi_5$	6	2	3	8	5	12	4	7	10	9	1	11
$\pi_6$	6	12	4	3	2	11	1	9	7	5	8	10
$\pi_7$	7	3	10	1	5	4	9	6	11	12	8	2
$\pi_8$	4	7	8	3	12	11	10	5	9	1	6	2
$\pi_9$	10	9	11	5	3	4	2	7	8	6	12	1
$\pi_{10}$	1	7	6	10	4	8	9	12	3	5	2	11

Fig. 1: Ten min-Hash permutations.

distance. Please explain in your report why it is efficient. In the attached “template/hamming.cpp” file, you are provided with a coding template.

### 3 Requirements

- Please write a report (in English) for the above two problems about how you address them and how well they work. Indicate your student number and student name in your PDF document and the email; For the first problem, please show the approximation accuracies in comparison to true *Jaccard* distance. For the second problem, please show the efficiency.
- Editing your document in *Latex* is recommended
- The deadline is 2025-Dec.-15.

### References

- [1] O. Chum, J. Philbin, M. Isard, A. Zisserman: Scalable Near Identical Image and Shot Detection, ACM CIVR 2007.