Similarity measures and KNN - Exercises
P. Rullo

Exercise 1. We are given the document/word matrix shown below, along with the following documents (represented as vectors)

$$X = (0,1,1,1,0) \text{ and } Y = (1,1,1,0,1)$$

Assign a label to both X and Y by using the Knn algorithm with k=3. Measure similarity by means of Cosine and Jaccard similarity metrics.

	parliament	italian	soccer	championship	govmt	class
d1	1	1	1	1	1	sport
d2	0	0	1	1	1	sport
d3	0	0	0	1	0	sport
d4	1	1	1	1	1	sport
d5	0	1	1	0	0	sport
d6	1	1	1	1	0	pilitics
d7	1	1	0	0	1	pilitics
d8	1	1	0	0	1	politics
d9	1	1	0	0	1	pilitics

Exercise 2. Consider the following objects

	A1	**A2**	**A3**	**A4**	**A5**
o1	A	PPP	0	0	X
o2	C	PPP	0	0	Y
o3	D	QQQ	1	1	Z

where

- A1 is ordinal with values (A,B,C,D)
- A2 is nominal
- A3 is asymmetric binary
- A4 is symmetric binary
- A5 is ordinal with values (X,Y,Z)

Say whether o2 is more similar to o1 or o3.