Solutions: Similarity and summary statistics

September 11, 2017

Task 1: Similarity

Compute the similarity between object 1 and objects 2–4 using the Jaccard coefficient and the cosine similarity.
(Note: for Jaccard similarity, all non-zero values should be treated as 1s.)

Object	attr 1	attr 2	attr 3	attr 4	attr 5	attr 6
1	1	2	0	0	0	0
2	1	0	1	1	1	0
3	3	0	0	0	0	3
4	2	4	0	0	0	0

Table 1: Sample data set.

Solution

Objects	**Jaccard**	**Cosine**
1 vs. 2	$\frac{1}{5}$	$\frac{1*1}{\sqrt{1^2+2^2}*\sqrt{1^2+1^2+1^2+1^2}} \approx 0.22$
1 vs. 3	$\frac{1}{3}$	$\frac{1*3}{\sqrt{1^2+2^2}*\sqrt{3^2+3^2}} \approx 0.32$
1 vs. 4	$\frac{2}{2} = 1$	$\frac{1*2+2*4}{\sqrt{1^2+2^2}*\sqrt{2^2+4^2}} = 1$

Table 2: Similarity between object 1 and objects 2–4.

Task 2: Summary statistics

Compute summary statistics for the following values: 17, 5, 3, 9, 49, 53, 11.
Round the results to two decimal places.

Solution

Values sorted: 3, 5, 9, 11, 17, 49, 53.

70^{th} **percentile**	49
Range	50
Median	11
Mean	21
Variance	441.33
Absolute Average Deviation	21.01

Table 3: Summary statistics.