

The background of the slide is a complex, abstract composition. It features a dark, reddish-brown base with a network of thin, light-colored lines forming a web-like structure. Scattered throughout are numerous small, colored dots in shades of green, blue, and orange. In the upper left, there's a horizontal band with a grid of small, light-colored plus signs. On the left side, there's a vertical strip containing a cluster of orange and red dots, with a horizontal bar of pink and white squares overlaid. The overall aesthetic is technical and data-driven.

Session 6: External Measures III: Pairwise Measures

In-video Quiz

❑ The following table summarizes the clustering result of two clustering algorithms, algorithm A and B.

❑ Which of the following statements are correct?

- 1) Based on Jaccard coefficient, Algorithm A is better;
- 2) Based on Rand Statistic, Algorithm A is better;
- 3) Fowlkes-Mallow Measure, Algorithm A is better;

❑ Answer: 1st and 3rd

❑ Explanation:

❑ For algorithm A, we have that

$$TP = \binom{8}{2} + \binom{3}{2} + \binom{6}{2} = 46$$

$$FN = \binom{9}{2} + \binom{9}{2} - TP = 26$$

$$FP = \binom{11}{2} + \binom{7}{2} - TP = 30$$

$$TN = \binom{18}{2} - TP - FN - FP = 51$$

❑ For algorithm B, we have that

$$TP = \binom{7}{2} + \binom{2}{2} + \binom{2}{2} + \binom{7}{2} = 44$$

$$FN = \binom{9}{2} + \binom{9}{2} - TP = 28$$

$$FP = \binom{9}{2} + \binom{9}{2} - TP = 28$$

$$TN = \binom{18}{2} - TP - FN - FP = 53$$

Algorithm A			
C\T	T ₁	T ₂	Sum
C ₁	8	3	11
C ₂	1	6	7
m _j	9	9	18

Algorithm B			
C\T	T ₁	T ₂	Sum
C ₁	7	2	9
C ₂	2	7	9
m _j	9	9	18

❑ For Jaccard coefficient

❑ Algorithm A: *Jaccard* = 0.451

❑ Algorithm B: *Jaccard* = 0.44

❑ Thus, Algorithm A is better

❑ For **Rand Statistic**:

❑ Algorithm A: 0.634

❑ Algorithm B: 0.634

❑ Thus, A and B are comparable;

❑ For **Fowlkes-Mallow Measure**

❑ Algorithm A: 0.622

❑ Algorithm B: 0.611

❑ Thus, Algorithm A is better