

## Purity-based evaluation for K-means

In this analysis, let  $k = r = 3$ . The contingency table is shown as follow:

NO.		T <sub>1</sub> iris-setosa	T <sub>2</sub> iris-versicolor	T <sub>3</sub> iris-virginica	n	Purity Score
1	C <sub>1</sub>	50	0	0	50	0.887
	C <sub>2</sub>	0	47	14	47	
	C <sub>3</sub>	0	3	36	36	
					133	
2	C <sub>1</sub>	0	2	36	36	0.893
	C <sub>2</sub>	50	0	0	50	
	C <sub>3</sub>	0	48	14	48	
					134	
3	C <sub>1</sub>	0	48	14	48	0.893
	C <sub>2</sub>	50	0	0	50	
	C <sub>3</sub>	0	2	36	36	
					134	
4	C <sub>1</sub>	50	0	0	50	0.893
	C <sub>2</sub>	0	2	36	36	
	C <sub>3</sub>	0	48	14	48	
					134	
5	C <sub>1</sub>	0	3	36	36	0.887
	C <sub>2</sub>	0	47	14	47	
	C <sub>3</sub>	50	0	0	50	
					133	
6	C <sub>1</sub>	0	48	14	48	0.893
	C <sub>2</sub>	50	0	0	50	
	C <sub>3</sub>	0	2	36	36	
					134	
7	C <sub>1</sub>	50	0	0	50	0.893
	C <sub>2</sub>	0	2	36	36	
	C <sub>3</sub>	0	48	14	48	
					134	
8	C <sub>1</sub>	0	3	36	36	0.887
	C <sub>2</sub>	0	47	14	47	
	C <sub>3</sub>	50	0	0	50	
					133	
9	C <sub>1</sub>	0	47	50	50	0.667
	C <sub>2</sub>	26	3	0	26	
	C <sub>3</sub>	24	0	0	24	
					100	
10	C <sub>1</sub>	0	48	14	48	0.893
	C <sub>2</sub>	0	2	36	36	
	C <sub>3</sub>	50	0	0	50	
					134	

The program is run for 10 times, and the best purity score is 0.893.