

## Homework 3 CS 5331: Data Analytics

-- Vibhuti Gupta & Gantaphon Chalumporn

**Q1)**

**(a)** Support of

$$s(\{e\}) = 8/10 = 0.8$$

$$s(\{b,d\}) = 2/10 = 0.2$$

$$s(\{b,d,e\}) = 2/10 = 0.2$$

**(b)** Confident of

$$c(\{b,d\} \rightarrow \{e\}) = 2/2 = 1.0$$

$$c(\{e\} \rightarrow \{b,d\}) = 2/8 = 0.25$$

The confident is NOT symmetric.

**(c)** When treating each customer ID as market basket, the support will be

$$s(\{e\}) = 4/5 = 0.8$$

$$s(\{b,d\}) = 2/5 = 0.4$$

$$s(\{b,d,e\}) = 2/5 = 0.4$$

**(d)** Use result from (c) to calculate confident

$$c(\{b,d\} \rightarrow \{e\}) = 2/2 = 1.0$$

$$c(\{e\} \rightarrow \{b,d\}) = 2/4 = 0.5$$

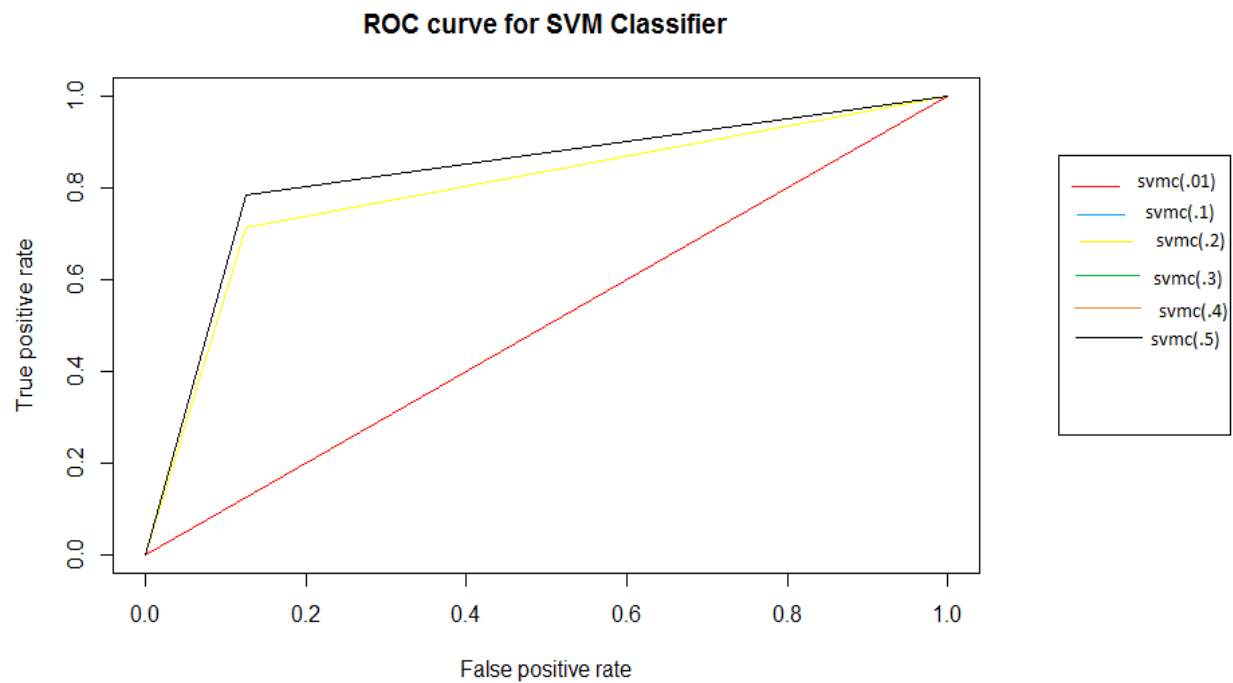
**(e)** No, there is no relationship between  $s_1 s_2$  or  $c_1 c_2$

**Q2) We used SVM to classified our Harvey Rescues Tweets**

**(1) Performance**

cost	Accuracy	Precision	Recall	F-measure
0.01	47	46.4	0	0
0.1	80	83.3	88	85.6
0.2	80	83.3	88	85.6
0.3	83	84.6	88	86.27
0.4	83	84.6	88	86.27
0.5	83	84.6	88	86.27

(2) Visualization of classification results, as shown in the figure. The our SVM classifier with cost equal to 0.3, 0.4, and 0.5 have the best performance.



(3) Compared with our best KNN Classifier from previous, the SVM has better performance in all aspects.

	Accuracy	Precision	Recall	F-measure
KNN (best)	50	60	86	70.6
SVM (best)	83	84.6	88	86.27