Birla Institute of Technology & Science, Pilani

Work Integrated Learning Programmes Division
Comprehensive Exam (Sample)

- Course Number: PCAMZC221
- Course Title: Unsupervised Learning and Association Rule Mining

---

1) Suppose you have been hired as a ML consultant for an Internet search engine company. Describe how techniques like clustering and anomaly/outlier detection can help the company by giving specific examples. [5 marks].

2) a) Given the Ratings of 5 employees on 2 different performance measurement indexes. Find clusters of similarly rated employees using K-means Algorithm, use Euclidean Distance as a distance measure and consider the number of clusters as 2. Choose "Stuart" and "Pavel" as initial centroids. [5 marks]

Employee Name	Rating from Index 1	Rating from Index 2
Jay	0	1
Stuart	3	0
Pavel	2	4
Andrey	2	1
Weifeng	3	5

b) K-Means results largely depends on the initialization of k-centroids, it may be optimal or suboptimal clustering depending on the choice of initial centroids. Suggest ways to overcome this issue. [3 marks]

3) Provide a clustering scenario where GMM using EM performs better than K-Means algorithm. Explain. [4 marks]

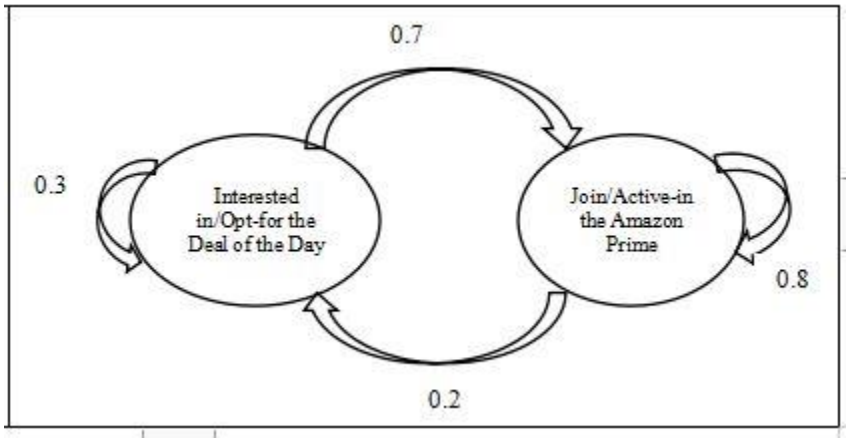
4) Following are the results of k-means (where k=3) clustering when applied on a labelled dataset that has 3 classes, class A, class B and class C.

		Actual			
		Class A	Class B	Class C	SUM
Clusters	Cluster 1	600	400	200	1200
	Cluster 2	1000	1200	200	2400
	Cluster 3	400	400	1600	2400
	SUM	2000	2000	2000	

Calculate F_1-Score for cluster 2 with respect to class B? [5 marks]

5) Refer to the markov model and find the probability that the e-commerce page usage pattern of the customer is the below order sequence observed. [2.5 Marks]

(Opt-for Deal , Join-Prime, Opt-for Deal, Opt-for Deal)



6) Marvel Entertainment would like to analyze the viewer's review comments of their 2018 film "Venom" to filter strong frequent association patterns among at least 50% of viewers. They are looking for only parameters in the below list and their intent of the analysis is to produce film under of similar genre for their next collaboration with Tencent Pictures. Representative sample reviews used for analysis is given below:

- Edit (or Editing or Edition or Scene or Visuals or Graphics)
- Expect(or Expectation)
- Hero
- Villain
- Anti-Hero,
- Action
- Comedy(or laugh or funny or humour)
- Plot (or Writing or story or character)

S.No	Ratings	Review Comments
1	1	Sadly, the writers went down the safer road and made the well known villain into a hero.
2	5	I love the anti-hero take on Venom. Funny action packed liked the story line. Not for kids but definitely recommend for a laugh
3	4	One of the best anti-hero character movie.Great display of acting and direction. The script is handled very well with good visuals. Yes, the villain could have been stronger.
4	3	how venom is a super villain but now they make him a hero and the bonding scene is really disappointing
5	4	First half goes very well in terms of setting the story and the venom universe, second part is more about action and there it has a few flaws but it is still fun and interesting to watch...but don't have too high expectations. Tom Hardy is in all the scene.
6	5	I like how they now thought something different from the comic storylines.. I personally loved the story plot!!

7	5	It's has funny moments which is necessary in movies like this. Not so serious or gravel voice like some action movies. Not so much fighting but more story which is a nice change.
8	5	the storyline was unique and i loved the way venom had a sense of humour to him throughout the whole movie.
9	1	The trailers were also poor editing, especially with the Wilhelm scream, so expectations were low. I was VERY excited at the beginning when I saw they were doing a Venom movie, but it's such a letdown. We're here for the villain/anti-hero black blob with a superiority complex.
10	3	Tho I dont call them plot holes. And its a real shame because it could of been a good movie.
11	4	It was entertaining, had plenty of action, and had a sprinkle on comedy at times. it was refreshing to see a movie where the villain is technically the good guy. Venom was really cool and the story was great
12	3	The character graph is given a heroic turnover . It starts slow and never picks up the pace. Both the lady and the villain needs more charisma .
13	1	Predictable plot, dialogue borrowed from a selection of scripts circa 1988. Receipts just confirm there's a sizeable audience with very low expectations

[2+3.5+1.5+1 = 8 Marks]

a. Convert the given data into transactional data.
b. Mine only the frequent 2-itemsets using Apriori algorithm.
c. Explain Apriori pruning principle in frequent items mining by using the results obtained from part b)
d. As per your understanding, justify in no more than 1 statement, why FP tree growth might be better suited for mining frequent itemsets instead of Apriori algorithm for this problem

7) Given the precomputed dissimilarity matrix of the movie production companies, apply the following clustering methods. Show stepwise working of the algorithm as discussed in class.

News Articles	A1	A2	A3	A4	A5	A6
A1	0	13	10	27	11	23
A2	13	0	5	31	9	18
A3	10	5	0	20	12	15
A4	27	31	20	0	8	21
A5	11	9	12	8	0	14
A6	23	18	15	21	14	0

a. Cluster the data using complete linkage algorithm and draw the dendrogram.

b. Explain in reference to the result of part a) What is the significance of cophenetic distance and where is it applied?

c. “It’s recommended to apply DBSCAN before hierarchically clustering with the complete linkage algorithm.” Justify this statement with appropriate situations in reference to the given sample problem. [4+2+1.5 = 7.5 Marks]

****