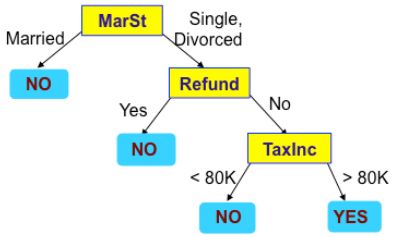
1. Given the following dataset:

   [1,2,3,4]
   [1,2,4]
   [1,2]
   [2,3,4]
   [2,3]
   [3,4]

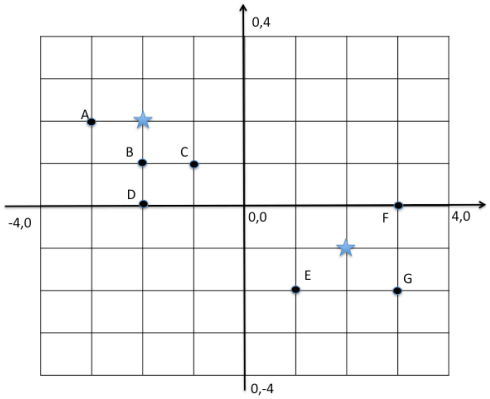
   emulate how Apriori works for detecting the frequent itemsets, with minimum support of 3.

2. Given the following decision tree and the test dataset, compute the *accuracy* and the *error rate.*

	categorical	categorical	continuous	class
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



3. Given these points and the two centroids (the stars), determine the two natural centroid-based clusters, using the Manhattan distance, and compute the SSE measure.



4. Discuss the following statement: DBSCAN is an algorithm that distinguishes between core, border, and noise objects.
5. What is the naive hypothesis at the basis of Bayesian classifiers? How do we train the model?
6. Precision and recall are two measures of quality/relevance of a set of results returned by a IR system. Discuss with an example.