

8. (4 points) Suppose we have a total of 100 items. The number of frequent items equals 10. How many candidate pairs will Apriori algorithm consider for counting in Pass 2?

Given  $n=100$  "10 choose 2"  $\rightarrow$

$$= \boxed{45}$$

9. (5 points) For Park-Chen-Yu (PCY) algorithm to improve over standard Apriori algorithm, the hash table in Pass 1 must eliminate at least the following fraction of candidate pairs

- a)  $1/3$
- b)  $1/2$
- ☒ c)  $2/3$
- d)  $3/4$

10. (5 points) The concept of *negative border* is relevant in the context of following algorithm:

a) Apriori

b) Park-Chen-Yu (PCY)

☒ c) Toivonen's Algorithm

☒ d) Savasere-Omiecinski-Navathe (SON) Algorithm

These algorithms are relevant with the negative border because they include a process of creating subsets or "chunks" from the data.

11. (5 points) When do you say an itemset  $I$  is in the *negative border*?

An itemset  $I$  is in the negative border if it is not deemed frequent in the sample however all its immediate subsets are.

### • Recommender Systems (20 points)

12. (5 points) What is the key idea behind content-based filtering algorithm to answer the basic filtering question: "will user  $U$  like item  $X$ ?"

- ☒ a) Look at what items  $U$  likes, and then check if  $X$  is similar to those items
- b) Look at which users like  $X$ , and then check if  $U$  is similar to those users

This is because content is between items not users.

13. (5 points) Root Mean Squared Error (RMSE) is the appropriate metric to evaluate the effectiveness or predictions of recommendation algorithms. (True/False)

True: The RMSE is used for comparison of predictions with known ratings. Thus something like a recommendation system can be evaluated from it.