

CptS 315 Introduction to Data Mining
Midterm Exam 1, Spring 2020
Exam date: Mar 12

Your Name and WSU ID: Zeid Al-Ameedi

Instructions.

- The maximum score of the exam is 100 points.
- Read all the questions before starting to answer. Try to answer those questions, which you think are easy from your perspective first. You can select a subset of questions that are worth 100 points to answer.
- Work efficiently. Most questions don't require much work. If you are spending more than 3 mins on any question, then you should try to re-think about it.
- Keep your answers short and simple.

• Short Questions (25 points)

Please keep your answers short (one sentence). For the True / False questions, please also provide a short justification.

0. (2 points) Incredibly hard question! What is the course number for this data mining class?

315

1. (5 points) Park-Chen-Yu (PCY) algorithm is always more efficient than Apriori algorithm for finding frequent item-pairs (True/False)

False: The PCY Algorithm must have 2/3 of the possible candidate pairs be eliminated in the first pass to be more efficient. The apriori (triangular method) will be 4 bytes/pair while PCY is 12 bytes/pair (tabular).

2. (5 points) User-user collaborative filtering and item-item collaborative filtering are duals of each other. Therefore, they both are equally accurate when applied to real-world applications. (True/False)

False: The Item-Item collaborative filtering essentially is better in implementation (or in practice).