Please complete the assigned problems to the best of your abilities. Ensure that the work you do is entirely your own, external resources are only used as permitted by the instructor, and all allowed sources are given proper credit for non-original content.

1 Recitation Problems

These problems are to be found in: **Mining of Massive Datasets, Online Edition** by *Jure Leskovec, Anand Rajaraman, Jeff Ullman.*

1.1 Chapter 9

Problems: 9.2.1,9.2.3,9.3.1,9.4.1

2 Practicum Problems

These problems will primarily reference the *lecture materials and the examples given in class* using **Orange** and **Python**. It is suggested that a *Jupyter/IPython* notebook be used for the programmatic components.

2.1 Problem 1

Load the *Movielens 100k* dataset (**ml-100k.zip**) into **Python** using Pandas dataframes. Build a user profile on *unscaled* data for both users **200** and **15**, and calculate the cosine similarity and distance between the user's preferences and the item/movie **95**. Which user would a recommender system suggest this movie to?

2.2 Problem 2

Load the *Movielens 100k* dataset (**ml-100k.zip**) into **Python** using Pandas dataframes. Convert the ratings data into a *utility matrix* representation, and find the **10** most similar users for user **1** based on cosine similarity of the user ratings data. Based on the average of of the ratings for item **508** from the similar users, what is the expected rating for this item for user **1**?