

CSC311 Final Report: Project Option 1

Leo Peckham, Longyue Wang, and Gursewak Sandhu

Part A

1. k -nearest neighbours

(a)

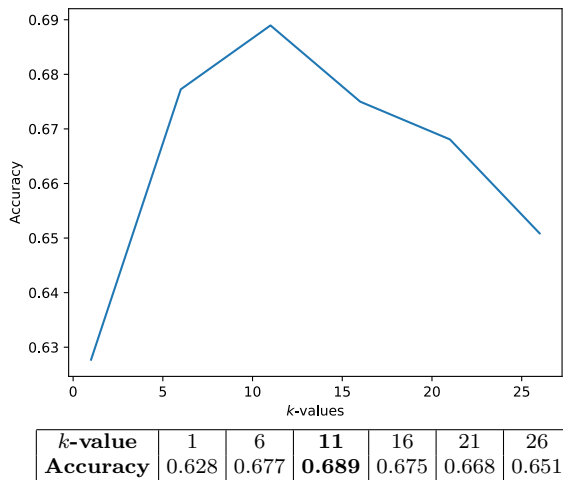


Figure 1: Accuracies of user-clustering KNN

(b) From this data, we can see that the k^* that gave us the best validation accuracy was $k^* = 11$. Running on test we achieve an accuracy of 0.683.

(c) Question-based clustering assumes that if two questions have very similar distributions of answers across the known students, then the questions will behave similarly for new students. Intuitively, if two questions ask about a specific theorem, then the same students who get the first one wrong because they forgot the statement of the theorem will get the second wrong as well.

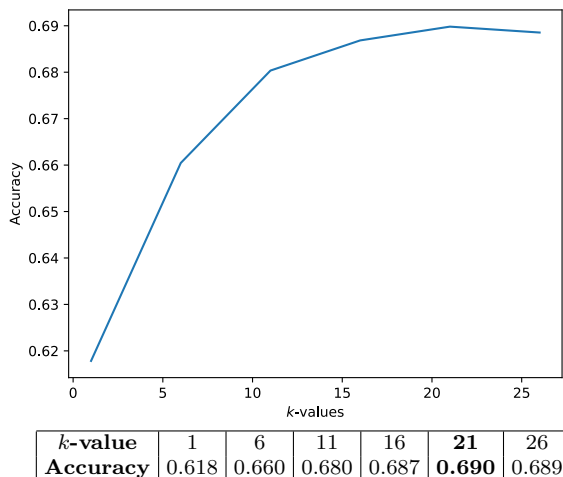


Figure 2: Accuracies of question-clustering KNN

From this data, we can see that the k^* that gave us the best validation accuracy was $k^* = 21$. Running on test we achieve an accuracy of 0.670.

(d) The performances across the board are similar. The user-based approach does better on test, but only by around a tenth of a percent accuracy. The graphs for different k -values do look quite different, though, with the user-clustering in figure 1 falling off much faster for high k -values than the question-clustering in figure 2.

(e) This method relies on a sufficient amount of data to be able to accurately cluster. If there is an example with a significant amount held-out, it will become inaccurate because many nearby users/questions will look similar.

The dimension of the features is also a problem. First, it can be extremely high, being either the number of students or the number of questions. This poses a problem for a KNN approach, especially one that uses a Euclidean distance like this one. Mapping to a smaller latent space first and using another distance function would help alleviate this problem. Second, if we add new students or questions to the dataset, the very *dimensionality* of our data will change. This makes it more difficult to improve our model with new data; mapping to a latent space would also fix this issue.

2. Item response theory

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3. Matrix factorization

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo,

lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4. Ensemble

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a

leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Part B

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.