K-Means Clustering in Academia

Rahul Rajeev

Bellevue University

DSC500-T302 Introduction to Data Science

Dr. Fadi Alsaleem

October 1, 2022

K-means clustering is an algorithm that determines which data points belong to each one of the k-clusters. These clusters contain points that have similar variance and minimize the within-cluster sum of squares (Bento, 2018). K-means clustering belongs to the algorithms used in unsupervised learning, where there are no explicit labels or classes for a group of data points. The model is responsible for classifying the data into implicit groups or clusters. Even though k-means clustering groups the data into the respective clusters, we are still responsible for determining the number of clusters required. The example that was covered in the article was regarding k-means clustering of data regarding workouts. A scenario where I can apply k-means clustering to is academic performance.

Schools are meant to provide high quality learning experiences with different curriculums and types of classes. Classes often range in difficulty and topics, and specialized classes in music and engineering could also be offered. The core of a student's performance is determined through a number of factors. Teachers monitor students' progress through the year by administering formative assessments, questioning, providing feedback, and progress reports (Goh, 2020). Formative assessments and personal questioning hold numerical graded scores, feedback could be on scales from 1-10, and progress reports are obviously a representation of the student's overall academic progress which could either be a qualitative or a quantitative variable. With these scores, it's probably a good idea to create a pair plot, similar to how it was done in the article. Once that's done, the "elbow method" can be used to determine the best value for k such that the clusters are not distorted (having only one value per cluster). This is where the heavy analysis starts to come into play.

K-means clustering can be used to determine the student's performance based on these factors and group them into the k-clusters. The k-clusters can then be used to identify students

who are at risk of failing academically. The clusters identify the areas where teachers may need to make adjustments and help each student accordingly. For example, if a student in a cluster is scoring high on formative assessments but is also falling behind on personal questioning, the teacher should work towards improving the student's confidence and slowly asking harder questions.

References

Bento, C. (2018, December 3). *K-means in real life: Clustering workout sessions*. Medium. Retrieved October 1, 2022, from https://towardsdatascience.com/k-means-in-real-life-clustering-workout-sessions-119946 9e8dd

Goh, Yann & Goh, Yeh & Chun Chieh, Yip & Ting, Chen & Bin, Raymond & Chen, Kah. (2020). Prediction of Students' Academic Performance by K-Means Clustering. Science Proceedings Series. 2. 1-6. 10.31580/sps.v2i1.1205.