KMEANS CLUSTERING

by

Yaser Alkayale

Submitted in partial fulfillment of the requirements
for the degree of Bachelor's of Computer Science, Honours

at

Dalhousie University
Halifax, Nova Scotia
April 2018

This thesis is in dedication to my grandfather whom I was named after.

Table of Contents

Abstract

Clustering is a well-known task that has been studied and used for decades. The idea is to take a set of items and group them into a number of clusters based on a similarity measure. K-means proposed in 1957 by Stuart Lloyd is one of the most widely used clustering algorithm and is still used today for its reasonably fast heuristic to find the clusters based on the Lloyd algorithm and more recent developments in that area. K-means has two main parts to clustering, the initial seeding process and the iteration process. The seeding process picks k initial seeds as cluster centres, and highly affects the accuracy of the final result in the algorithm. The iteration process dominates running time to move the centres around until it converges to an optimum. In this paper, we discuss a new method of the seeding process that gives us more accurate seeds to start the algorithm. We also discuss a novel approach to find an approximation of the correct number of clusters for a given dataset.

Acknowledgements

Thanks to all the little people who make me look tall.

Chapter 1

Introduction

Get it done! Use reference material by Limpet [2] or Gooses, Mittelback, and Samarin [1].

To test if the margins are satisfactory, let us generate a lot of garbage text: This sentence goes on, and and here we should be around top of the page 2, and we go on, and on, and on, and on,

and on. This following line ______________ should be exactly 5cm long. It can be used to check the typesetting process. And now we go on, and on,

and on, and on.

Chapter 2

Doing It

2.1 Getting Ready

Get all the parts that I need. I can throw in a whole pile of terms like preparation, methodology, forethought, and analysis as examples for me to use in the future.

2.2 Next Step

Do it!

Of course, you have to have pictures to show how you did it to make people understand things better.

Chapter 3

Conclusion

Did it![2]

Bibliography

[1] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, 1994.

[2] Leslie Lamport. *A Document Preparation System Latex User's Guide and Reference Manual*. Addison-Wesley, 1986.