B.Sc. in Computer Science and Engineering Thesis

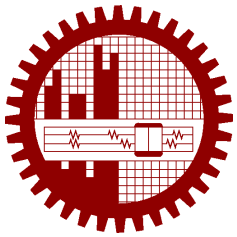
Selection of K in K-Means Algorithm

Submitted by

MD. Nayeem Reza
201205106

Supervised by

Dr. Md. Monirul Islam



Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

September 2017

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, "Selection of K in K-Means Algorithm", is the outcome of the investigation and research carried out by us under the supervision of Dr. Md. Monirul Islam.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

MD. Nayeem Reza
201205106

CERTIFICATION

This thesis titled, **"Selection of K in K-Means Algorithm"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in September 2017.

Group Members:

MD. Nayeem Reza

Supervisor:

Dr. Md. Monirul Islam
Professor
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

First of all I would like to thank my supervisor, Dr. Md. Monirul Islam for introducing me to the amazingly interesting world of Machine Learning and Data Analysis. And he is the very first person who taught me how to perform research work efficiently. Without him and his continuous supervision, guidance and valuable advice, it would have been impossible for me to come at this point and have some output from the thesis. I am specially grateful to him for allowing me greater freedom in choosing the problems to work on, for his encouragement at times of disappointment, and for his patience with my wildly sporadic work habits.

I would also want to thank my fellow friends Md. Thohidul Islam, Shariful Islam Foysal and Md. Moyeen Uddin Fahim for their encouragement, insightful comments, and hard questions. They gave their best support to complete my thesis timely. I would like to express my gratitude to all our teachers. Their motivation and encouragement in addition to the education they provided meant a lot to me.

Last but not the least, I would like to thank my family, my parents, for giving birth to me at the first place and supporting me spiritually throughout my life.

Dhaka
September 2017

MD. Nayeem Reza

Contents

CANDIDATES' DECLARATION i

CERTIFICATION ii

ACKNOWLEDGEMENT iii

List of Figures vi

List of Tables vii

List of Algorithms viii

ABSTRACT ix

1 Introduction 1

1.1 Introduction . . . 1

1.1.1 Clustering . . . 1

1.1.2 Goals . . . 2

1.1.3 Applications . . . 2

1.1.4 Requirements . . . 2

1.1.5 Problems . . . 3

1.1.6 Classification . . . 3

1.2 Literature Survey . . . 6

1.2.1 Distance Measure . . . 6

1.3 Research Gap . . . 7

1.4 Objective . . . 7

1.5 Thesis Organization . . . 7

1.6 Cross Referencing . . . 7

1.7 How to Write a Section . . . 7

1.8 How to Add Table and Figures . . . 7

2 Background 10

2.1 Introduction . . . 10

2.1.1 Machine Learning . . . 10

2.2 A Deep Dive into Cluster Analysis . . . 12
2.3 A . . . 12
2.4 B . . . 12
2.5 C . . . 12
2.6 D . . . 12

3 Citation Examples **13**
3.1 See the Citations . . . 13

4 Another Chapter **14**
4.1 A Section . . . 14
4.1.1 This is a Subsection . . . 14
4.2 And Another Section . . . 14

5 Index Creation **17**
5.1 BUET . . . 17
5.2 Campus . . . 17
5.3 History . . . 17
5.4 Students . . . 18
5.5 Departments . . . 18

6 k-safe Labeling of Petersen Graph **19**

References **20**

Index **23**

A Algorithms **24**
A.1 Sample Algorithm . . . 24

B Codes **25**
B.1 Sample Code . . . 25
B.2 Another Sample Code . . . 26

List of Figures

1.1 Clustering Process 1

List of Tables

1.1 Performance table of *Block reversal* in a heuristic algorithm 8

List of Algorithms

1 Calculate $y = x^n$ 24

ABSTRACT

Write your thesis abstract here.

Chapter 1

Introduction

1.1 Introduction

1.1.1 Clustering

Clustering can be considered as the most important *unsupervised learning* problem; so, as every other problem of this type, it deals with finding a *structure* in a collection of given unlabeled datasets. A very common informal definition of clustering could be "the process of organizing objects into groups whose members are similar in some features of the given dataset". A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

We can show this with a simple graphical example:

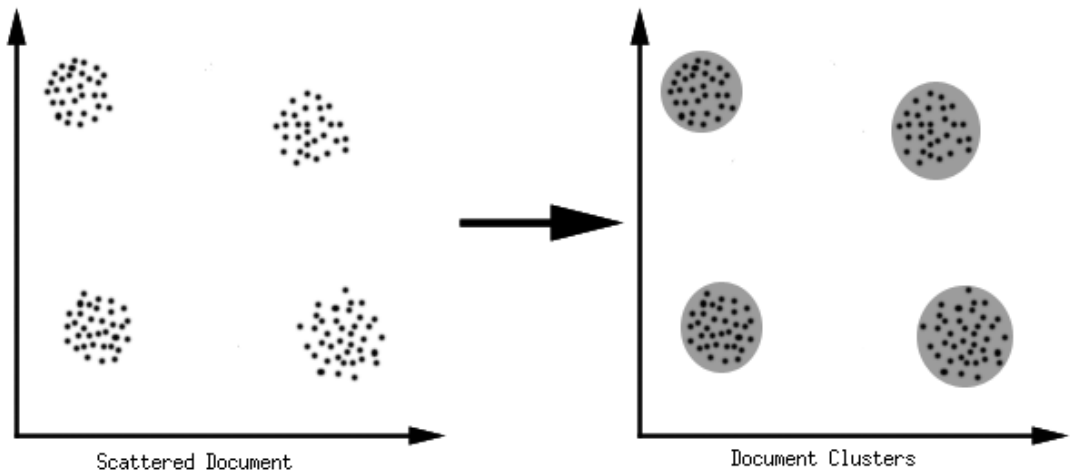


Figure 1.1: Clustering Process

In Figure 1.1 we can easily identify the 4 clusters into which the data can be divided; the

similarity criterion is *distance*: two or more objects belong to the same cluster if they are "close" according to a given distance (in this case geometrical distance). This is called *distance-based clustering*. Another kind of clustering is *conceptual clustering*: two or more objects belong to the same cluster if this one defines a concept common to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.

1.1.2 Goals

The main goal of clustering is to determine the intrinsic grouping in a given set of unlabeled data. But there is no such factors to decide what constitutes a good clustering. It can be shown that there is no absolute "best" criterion which would be independent of the final aim of the clustering. Consequently, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), in finding "natural clusters" and describe their unknown properties ("natural" data types), in finding useful and suitable groupings ("useful" data classes) or in finding unusual data objects (outlier detection).

1.1.3 Applications

Clustering algorithms can be applied in a very wide range of fields. In *Marketing* we can use clustering to find groups of customers with similar behavior from a large database of customers data containing their properties and past buying records. In *Biology* we can classify plants and animals by their given feature sets. Book ordering and sorting can be a good application of clustering in the *Libraries*. *Insurance* companies use clustering for identifying groups of insurance policy holders with a high average claim cost. Recently urban developers are using clustering methods for identifying groups of houses according to their house type, value and geographical location in *city-planning*. To identify dangerous zones of earthquake we can observe earthquake epicenters by using clustering algorithms on. These days clustering algorithms are mostly used in *WWW* for document classification; clustering weblog data to discover groups of similar access patterns.

1.1.4 Requirements

The main requirements that a clustering algorithm should satisfy are:

- **Scalability** : We need highly scalable clustering algorithms to deal with large databases.

- **Ability to deal with different kinds of attributes** : Algorithms should be capable to be applied on any kind of data such as interval-based (numerical) data, categorical, and binary data.
- **Discovery of clusters with attribute shape** : The clustering algorithm should be capable of detecting clusters of arbitrary shape. They should not be bounded to only distance measures that tend to find spherical cluster of small sizes.
- **High dimensionality** : The clustering algorithm should not only be able to handle low-dimensional data but also the high dimensional space.
- **Ability to deal with noisy data** : Databases contain noisy, missing or erroneous data. Some algorithms are sensitive to such data and may lead to poor quality clusters.
- **Interpretability** : The clustering results should be interpretable, comprehensible, and usable.

1.1.5 Problems

There are a number of problems with clustering. Among them:

- Current clustering techniques do not address all the requirements adequately (and concurrently);
- Dealing with large number of dimensions and large number of data items can be problematic because of time complexity;
- The effectiveness of the method depends on the definition of distance (for distance-based clustering);
- If an obvious distance measure doesnt exist we must define it, which is not always easy, especially in multi-dimensional spaces;
- The result of the clustering algorithm (that in many cases can be arbitrary itself) can be interpreted in different ways.

1.1.6 Classification

Clustering algorithms may be classified as listed below:

- **Hierarchical Clustering** : Hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. These algorithms connect

"objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a *dendrogram*, which explains where the common name "Hierarchical Clustering" comes from: these algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. In a dendrogram, the Y-axis marks the distance at which the clusters merge, while the objects are placed along the X-axis such that the clusters don't mix.

Connectivity based clustering is a whole family of methods that differ by the way distances are computed. Apart from the usual choice of *distance functions*, the user also needs to decide on the linkage criterion (since a cluster consists of multiple objects, there are multiple candidates to compute the distance to) to use. Popular choices are known as *single-linkage clustering* (the minimum of object distances), *complete linkage clustering* (the maximum of object distances) or UPGMA ("Unweighted Pair Group Method with Arithmetic Mean", also known as average linkage clustering). Furthermore, hierarchical clustering can be agglomerative (starting with single elements and aggregating them into clusters) or divisive (starting with the complete data set and dividing it into partitions).

These methods will not produce a unique partitioning of the data set, but a hierarchy from which the user still needs to choose appropriate clusters. They are not very robust towards outliers, which will either show up as additional clusters or even cause other clusters to merge (known as "chaining phenomenon", in particular with single-linkage clustering). In the general case, the complexity is $O(n^3)$ for agglomerative clustering and $O(2^{n-1})$ for *divisive clustering*, which makes them too slow for large data sets. For some special cases, optimal efficient methods (of complexity $O(n^2)$ are known as single-linkage and complete-linkage clustering. In the *data mining* community these methods are recognized as a theoretical foundation of cluster analysis, but often considered obsolete. They did however provide inspiration for many later methods such as density based clustering.

- **Centroid-based Clustering** : In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k, k-means clustering gives a formal definition as an optimization problem: find the k cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized.

  The optimization problem itself is known to be NP-hard, and thus the common approach is to search only for approximate solutions. A particularly well known approximative method is Lloyd's algorithm, often actually referred to as "k-means algorithm". It does however only find a local optimum, and is commonly run multiple times with different random initializations. Variations of k-means often include such optimizations as choosing the best of multiple runs, but also restricting the centroids to members of the data set

(k-medoids), choosing medians (k-medians clustering), choosing the initial centers less randomly (k-means++) or allowing a fuzzy cluster assignment (fuzzy c-means).

Most k-means-type algorithms require the number of clusters - k to be specified in advance, which is considered to be one of the biggest drawbacks of these algorithms. Furthermore, the algorithms prefer clusters of approximately similar size, as they will always assign an object to the nearest centroid. This often leads to incorrectly cut borders of clusters (which is not surprising since the algorithm optimizes cluster centers, not cluster borders).

- **Distribution-based clustering** : The clustering model most closely related to statistics is based on distribution models. Clusters can then easily be defined as objects belonging most likely to the same distribution. A convenient property of this approach is that this closely resembles the way artificial data sets are generated: by sampling random objects from a distribution.

  While the theoretical foundation of these methods is excellent, they suffer from one key problem known as overfitting, unless constraints are put on the model complexity. A more complex model will usually be able to explain the data better, which makes choosing the appropriate model complexity inherently difficult.

  One prominent method is known as Gaussian mixture models (using the expectation-maximization algorithm). Here, the data set is usually modelled with a fixed (to avoid overfitting) number of Gaussian distributions that are initialized randomly and whose parameters are iteratively optimized to better fit the data set. This will converge to a local optimum, so multiple runs may produce different results. In order to obtain a hard clustering, objects are often then assigned to the Gaussian distribution they most likely belong to; for soft clusterings, this is not necessary.

  Distribution-based clustering produces complex models for clusters that can capture correlation and dependence between attributes. However, these algorithms put an extra burden on the user: for many real data sets, there may be no concisely defined mathematical model (e.g. assuming Gaussian distributions is a rather strong assumption on the data).

- **Density-based Clustering** : In density-based clustering, clusters are defined as areas of higher density than the remainder of the data set. Objects in these sparse areas - that are required to separate clusters - are usually considered to be noise and border points.

  The most popular density based clustering method is DBSCAN. In contrast to many newer methods, it features a well-defined cluster model called “density-reachability”. Similar to linkage based clustering, it is based on connecting points within certain distance thresholds. However, it only connects points that satisfy a density criterion, in the original variant defined as a minimum number of other objects within this radius. A cluster consists of all density-connected objects (which can form a cluster of an arbitrary

shape, in contrast to many other methods) plus all objects that are within these objects' range. Another interesting property of DBSCAN is that its complexity is fairly low - it requires a linear number of range queries on the database - and that it will discover essentially the same results (it is deterministic for core and noise points, but not for border points) in each run, therefore there is no need to run it multiple times. OPTICS is a generalization of DBSCAN that removes the need to choose an appropriate value for the range parameter ε, and produces a hierarchical result related to that of linkage clustering. DeLi-Clu, Density-Link-Clustering combines ideas from single-linkage clustering and OPTICS, eliminating the ε parameter entirely and offering performance improvements over OPTICS by using an R-tree index.

The key drawback of DBSCAN and OPTICS is that they expect some kind of density drop to detect cluster borders. On data sets with, for example, overlapping Gaussian distributions - a common use case in artificial data - the cluster borders produced by these algorithms will often look arbitrary, because the cluster density decreases continuously. On a data set consisting of mixtures of Gaussians, these algorithms are nearly always outperformed by methods such as EM clustering that are able to precisely model this kind of data.

Mean-shift is a clustering approach where each object is moved to the densest area in its vicinity, based on kernel density estimation. Eventually, objects converge to local maxima of density. Similar to k-means clustering, these "density attractors" can serve as representatives for the data set, but mean-shift can detect arbitrary-shaped clusters similar to DBSCAN. Due to the expensive iterative procedure and density estimation, mean-shift is usually slower than DBSCAN or k-Means. Besides that, the applicability of the mean-shift algorithm to multidimensional data is hindered by the unsmooth behaviour of the kernel density estimate, which results in over-fragmentation of cluster tails.

1.2 Literature Survey

1.2.1 Distance Measure

An important component of a clustering algorithm is the distance measure between data points. If the components of the data instance vectors are all in the same physical units then it is possible that the simple Euclidean distance metric is sufficient to successfully group similar data instances. However, even in this case the Euclidean distance can sometimes be misleading. Figure shown below illustrates this with an example of the width and height measurements of an object. Despite both measurements being taken in the same physical units, an informed decision has to be made as to the relative scaling. As the figure shows, different scalings can lead

to different clusterings. Notice however that this is not only a graphic issue: the problem arises from the mathematical formula used to combine the distances between the single components of the data feature vectors into a unique distance measure that can be used for clustering purposes: different formulas leads to different clusterings. Again, domain knowledge must be used to guide the formulation of a suitable distance measure for each particular application.

1.3 Research Gap

1.4 Objective

1.5 Thesis Organization

1.6 Cross Referencing

We have incorporated the `\cref` or `\Cref` command from `cleveref` package in this system. This will automatically insert words like Figure, Table etc. in your text.

See these examples:

- **??** is a sample figure.
- Table 1.1 is a table.
- Section 3.1 in Chapter 3 shows some examples of citations.

1.7 How to Write a Section

This is for writing section.

1.8 How to Add Table and Figures

You should refer a figure as, "**??** is a sample figure".

Then we applied same test cases to our modified algorithm i.e. the heuristic algorithm with our new operation *Block Reversal*. The performance is shown in Table 1.1.

These are some dummy text used as page fillers only.

Table 1.1: Performance table of *Block reversal* in a heuristic algorithm

α	αn	Test Cases											Average # of calculated operation
		1	2	3	4	5	6	7	8	9	10	11	
0.1	2	2	2	2	2	2	2	2	2	2	2	2	2
0.2	4	4	4	5	2	4	4	4	4	2	4	4	3.73
0.3	6	5	6	6	6	6	7	6	5	6	6	6	5.91
0.4	8	7	8	5	6	7	6	6	7	8	8	7	6.82
0.5	10	9	10	6	12	10	8	10	10	7	7	10	9
0.6	12	9	12	16	10	12	12	9	11	12	9	12	11.27
0.7	14	13	7	18	15	14	8	13	11	13	13	14	12.64
0.8	16	10	17	14	16	13	16	13	11	13	17	13	13.91
0.9	18	14	16	15	12	15	11	15	11	15	12	12	13.45
1	20	18	11	13	11	13	15	17	17	13	18	12	14.36

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras et ultricies massa. Nulla a sapien lobortis, dignissim nibh in, aliquet mauris. Integer at dictum metus. Quisque in tortor congue ipsum ultricies tristique. Maecenas ut tortor dapibus, sagittis enim at, tincidunt massa. Ut sollicitudin sagittis ipsum, ac tincidunt quam gravida ac. Nullam quis faucibus purus. Aliquam vel pretium turpis. Aliquam a quam non ex interdum sagittis id vitae quam. Nullam sodales ligula malesuada maximus consequat. Proin a justo eget lacus vulputate maximus luctus vitae enim. Aliquam libero turpis, pharetra a tincidunt ac, pulvinar sit amet urna. Pellentesque eget rutrum diam, in faucibus sapien. Aenean sit amet est felis. Aliquam dolor eros, porttitor quis volutpat eget, posuere a ligula. Proin id velit ac lorem finibus pellentesque.

Maecenas vitae interdum mi. Aenean commodo nisl massa, at pharetra libero cursus vitae. In hac habitasse platea dictumst. Suspendisse iaculis euismod dui, et cursus diam. Nullam euismod, est ut dapibus condimentum, lorem eros suscipit risus, sit amet hendrerit justo tortor nec lorem. Morbi et mi eget erat bibendum porta. Ut tristique ultricies commodo. Nullam iaculis ligula sed lacinia ornare.

Sed ultricies cursus nisi at vestibulum. Aenean laoreet viverra efficitur. Ut eget sapien lorem. Mauris malesuada, augue in pulvinar consectetur, ex tortor tristique ligula, sit amet faucibus metus lectus interdum nisl. Nam eget turpis vitae ligula pulvinar bibendum a ut ipsum. Mauris fringilla lacinia malesuada. Fusce id orci velit. Donec tristique rhoncus urna, a hendrerit arcu vehicula imperdiet. Integer tristique erat at gravida condimentum. Sed ornare cursus quam, eget tincidunt enim bibendum sed. Aliquam elementum ligula scelerisque leo sagittis, quis convallis elit dictum. Donec sit amet orci aliquam, ultricies sapien nec, gravida nisi. Etiam et pulvinar diam, et pellentesque arcu. Nulla interdum metus sed aliquet consequat.

Proin in mi id nulla interdum aliquet ac quis arcu. Duis blandit sapien commodo turpis hendrerit pharetra. Phasellus sit amet justo orci. Proin mattis nisl dictum viverra fringilla. Interdum et malesuada fames ac ante ipsum primis in faucibus. Curabitur facilisis euismod augue vestibu-

lum tincidunt. Nullam nulla quam, volutpat vitae efficitur eget, porta sit amet nunc. Phasellus pharetra est eget urna ornare volutpat. Aenean ultrices, libero eget porttitor fringilla, purus tortor accumsan neque, sit amet viverra felis tortor eget justo. Nunc id metus a purus tempus euismod condimentum non lacus. Nam vitae diam aliquam, facilisis diam quis, pharetra nunc. Nulla eget vestibulum tellus, ut cursus tellus. Vestibulum euismod pellentesque sodales.

Maecenas at mi interdum, faucibus lorem sed, hendrerit nisi. In vitae augue consequat diam commodo porta sit amet eu purus. Mauris mattis condimentum feugiat. Nulla commodo molestie risus vitae maximus. Proin hendrerit neque malesuada urna laoreet convallis. Etiam a diam pulvinar, auctor sem ac, hendrerit risus. Ut urna urna, venenatis ac tellus non, scelerisque tristique ligula. Vestibulum sollicitudin vel leo malesuada accumsan. Donec sit amet erat diam. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Vivamus odio dui, scelerisque et lorem egestas, posuere ullamcorper nunc. Integer varius nunc nec velit tincidunt commodo. Mauris rhoncus ultrices sapien non suscipit.

End of dummy text.

Chapter 2

Background

2.1 Introduction

2.1.1 Machine Learning

Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data such algorithms overcome following strictly static program instructions by making data driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible; example applications include spam filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), search engines and computer vision.

Tom M. Mitchell [1] provided a widely quoted, more formal definition:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning "signal" or "feedback" available to a learning system. These are:

Supervised Learning
Supervised learning is where you have input variables (x) and an output variable (Y) and you

use an algorithm to learn the mapping function from the input to the output. Exact functon may be as $Y = f(X)$. The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance. Supervised learning problems can be further grouped into regression and classification problems.

- **Classification** : A classification problem is when the output variable is a category, such as "red" or "blue" or "disease" and "no disease".
- **Regression** : A regression problem is when the output variable is a real value, such as "dollars" or "weight".

Unsupervised Learning

No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).Unsupervised learning is where you only have input data (X) and no corresponding output variables.

The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data.Unsupervised learning problems can be further grouped into clustering and association problems.

- **Clustering** : A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association** : An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Semi-Supervised Machine Learning

Problems where there is a large amount of input data (X) and only some of the data is labeled (Y) are called semi-supervised learning problems.These problems sit in between both supervised and unsupervised learning. A good example is a photo archive where only some of the images are labeled, (e.g. dog, cat, person) and the majority are unlabeled.

Many real world machine learning problems fall into this area. This is because it can be expensive or time-consuming to label data as it may require access to domain experts. Whereas unlabeled data is cheap and easy to collect and store.

Reinforcement Learning
A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). The program is provided feedback in terms of rewards and punishments as it navigates its problem space.

2.2 A Deep Dive into Cluster Analysis

Cluster analysis groups data objects based only on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better or more distinct the clustering.

Cluster analysis is related to other techniques that are used to divide data objects into groups. For instance, clustering can be regarded as a form of classification in that it creates a labeling of objects with class (cluster) labels. However, it derives these labels only from the data. In contrast, classification in the sense of our previous section is **supervised classification**; i.e., new, unlabeled objects are assigned a class label using a model developed from objects with known class labels. For this reason, cluster analysis is sometimes referred to as **unsupervised classification**. When the term classification is used without any qualification within data mining, it typically refers to supervised classification.

2.3 A

2.4 B

2.5 C

2.6 D

Chapter 3

Citation Examples

In this chapter we show how we can cite the references.

3.1 See the Citations

As discussed by authors in [2–4] we can further show how this affects us. Moreover [5–12] can be examples for the previous works. Among these [11, 13–18] are the prominent ones. Also you can take a look at [19–26].

Chapter 4

Another Chapter

4.1 A Section

Some text.

4.1.1 This is a Subsection

And some more.

This is a Subsubsection

Yet some more.

4.2 And Another Section

Here are some dummy texts.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras et ultricies massa. Nulla a sapien lobortis, dignissim nibh in, aliquet mauris. Integer at dictum metus. Quisque in tortor congue ipsum ultricies tristique. Maecenas ut tortor dapibus, sagittis enim at, tincidunt massa. Ut sollicitudin sagittis ipsum, ac tincidunt quam gravida ac. Nullam quis faucibus purus. Aliquam vel pretium turpis. Aliquam a quam non ex interdum sagittis id vitae quam. Nullam sodales ligula malesuada maximus consequat. Proin a justo eget lacus vulputate maximus luctus vitae enim. Aliquam libero turpis, pharetra a tincidunt ac, pulvinar sit amet urna. Pellentesque eget rutrum diam, in faucibus sapien. Aenean sit amet est felis. Aliquam dolor eros, porttitor quis volutpat eget, posuere a ligula. Proin id velit ac lorem finibus pellentesque.

Maecenas vitae interdum mi. Aenean commodo nisl massa, at pharetra libero cursus vitae. In hac habitasse platea dictumst. Suspendisse iaculis euismod dui, et cursus diam. Nullam euismod, est ut dapibus condimentum, lorem eros suscipit risus, sit amet hendrerit justo tortor nec lorem. Morbi et mi eget erat bibendum porta. Ut tristique ultricies commodo. Nullam iaculis ligula sed lacinia ornare.

Sed ultricies cursus nisi at vestibulum. Aenean laoreet viverra efficitur. Ut eget sapien lorem. Mauris malesuada, augue in pulvinar consectetur, ex tortor tristique ligula, sit amet faucibus metus lectus interdum nisl. Nam eget turpis vitae ligula pulvinar bibendum a ut ipsum. Mauris fringilla lacinia malesuada. Fusce id orci velit. Donec tristique rhoncus urna, a hendrerit arcu vehicula imperdiet. Integer tristique erat at gravida condimentum. Sed ornare cursus quam, eget tincidunt enim bibendum sed. Aliquam elementum ligula scelerisque leo sagittis, quis convallis elit dictum. Donec sit amet orci aliquam, ultricies sapien nec, gravida nisi. Etiam et pulvinar diam, et pellentesque arcu. Nulla interdum metus sed aliquet consequat.

Proin in mi id nulla interdum aliquet ac quis arcu. Duis blandit sapien commodo turpis hendrerit pharetra. Phasellus sit amet justo orci. Proin mattis nisl dictum viverra fringilla. Interdum et malesuada fames ac ante ipsum primis in faucibus. Curabitur facilisis euismod augue vestibulum tincidunt. Nullam nulla quam, volutpat vitae efficitur eget, porta sit amet nunc. Phasellus pharetra est eget urna ornare volutpat. Aenean ultrices, libero eget porttitor fringilla, purus tortor accumsan neque, sit amet viverra felis tortor eget justo. Nunc id metus a purus tempus euismod condimentum non lacus. Nam vitae diam aliquam, facilisis diam quis, pharetra nunc. Nulla eget vestibulum tellus, ut cursus tellus. Vestibulum euismod pellentesque sodales.

Maecenas at mi interdum, faucibus lorem sed, hendrerit nisi. In vitae augue consequat diam commodo porta sit amet eu purus. Mauris mattis condimentum feugiat. Nulla commodo molestie risus vitae maximus. Proin hendrerit neque malesuada urna laoreet convallis. Etiam a diam pulvinar, auctor sem ac, hendrerit risus. Ut urna urna, venenatis ac tellus non, scelerisque tristique ligula. Vestibulum sollicitudin vel leo malesuada accumsan. Donec sit amet erat diam. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Vivamus odio dui, scelerisque et lorem egestas, posuere ullamcorper nunc. Integer varius nunc nec velit tincidunt commodo. Mauris rhoncus ultrices sapien non suscipit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras et ultricies massa. Nulla a sapien lobortis, dignissim nibh in, aliquet mauris. Integer at dictum metus. Quisque in tortor congue ipsum ultricies tristique. Maecenas ut tortor dapibus, sagittis enim at, tincidunt massa. Ut sollicitudin sagittis ipsum, ac tincidunt quam gravida ac. Nullam quis faucibus purus. Aliquam vel pretium turpis. Aliquam a quam non ex interdum sagittis id vitae quam. Nullam sodales ligula malesuada maximus consequat. Proin a justo eget lacus vulputate maximus luctus vitae enim. Aliquam libero turpis, pharetra a tincidunt ac, pulvinar sit amet urna. Pellentesque eget rutrum diam, in faucibus sapien. Aenean sit amet est felis. Aliquam dolor eros, porttitor quis

volutpat eget, posuere a ligula. Proin id velit ac lorem finibus pellentesque.

Maecenas vitae interdum mi. Aenean commodo nisl massa, at pharetra libero cursus vitae. In hac habitasse platea dictumst. Suspendisse iaculis euismod dui, et cursus diam. Nullam euismod, est ut dapibus condimentum, lorem eros suscipit risus, sit amet hendrerit justo tortor nec lorem. Morbi et mi eget erat bibendum porta. Ut tristique ultricies commodo. Nullam iaculis ligula sed lacinia ornare.

Sed ultricies cursus nisi at vestibulum. Aenean laoreet viverra efficitur. Ut eget sapien lorem. Mauris malesuada, augue in pulvinar consectetur, ex tortor tristique ligula, sit amet faucibus metus lectus interdum nisl. Nam eget turpis vitae ligula pulvinar bibendum a ut ipsum. Mauris fringilla lacinia malesuada. Fusce id orci velit. Donec tristique rhoncus urna, a hendrerit arcu vehicula imperdiet. Integer tristique erat at gravida condimentum. Sed ornare cursus quam, eget tincidunt enim bibendum sed. Aliquam elementum ligula scelerisque leo sagittis, quis convallis elit dictum. Donec sit amet orci aliquam, ultricies sapien nec, gravida nisi. Etiam et pulvinar diam, et pellentesque arcu. Nulla interdum metus sed aliquet consequat.

Proin in mi id nulla interdum aliquet ac quis arcu. Duis blandit sapien commodo turpis hendrerit pharetra. Phasellus sit amet justo orci. Proin mattis nisl dictum viverra fringilla. Interdum et malesuada fames ac ante ipsum primis in faucibus. Curabitur facilisis euismod augue vestibulum tincidunt. Nullam nulla quam, volutpat vitae efficitur eget, porta sit amet nunc. Phasellus pharetra est eget urna ornare volutpat. Aenean ultrices, libero eget porttitor fringilla, purus tortor accumsan neque, sit amet viverra felis tortor eget justo. Nunc id metus a purus tempus euismod condimentum non lacus. Nam vitae diam aliquam, facilisis diam quis, pharetra nunc. Nulla eget vestibulum tellus, ut cursus tellus. Vestibulum euismod pellentesque sodales.

Maecenas at mi interdum, faucibus lorem sed, hendrerit nisi. In vitae augue consequat diam commodo porta sit amet eu purus. Mauris mattis condimentum feugiat. Nulla commodo molestie risus vitae maximus. Proin hendrerit neque malesuada urna laoreet convallis. Etiam a diam pulvinar, auctor sem ac, hendrerit risus. Ut urna urna, venenatis ac tellus non, scelerisque tristique ligula. Vestibulum sollicitudin vel leo malesuada accumsan. Donec sit amet erat diam. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Vivamus odio dui, scelerisque et lorem egestas, posuere ullamcorper nunc. Integer varius nunc nec velit tincidunt commodo. Mauris rhoncus ultrices sapien non suscipit.

Chapter 5

Index Creation

5.1 BUET

Bangladesh University of Engineering and Technology, abbreviated as BUET, is one of the most prestigious institutions for higher studies in the country. About 5500 students are pursuing undergraduate and postgraduate studies in engineering, architecture, planning and science in this institution. At present, BUET has sixteen teaching departments under five faculties and it has three institutes. Every year the intake of undergraduate students is around 900, while the intake of graduate students in Master's and PhD programs is around 1000. A total of about five hundred teachers are teaching in these departments and institutes. There are additional teaching posts like Dr. Rashid Professor, Professor Emeritus and Supernumerary Professors.

5.2 Campus

The BUET campus is in the heart of Dhaka — the capital city of Bangladesh. It has a compact campus with halls of residence within walking distances of the academic buildings. The physical expansion of the University over the last three decades has been impressive with construction of new academic buildings, auditorium complex, halls of residence, etc.

5.3 History

BUET is the oldest institution for the study of Engineering and Architecture in Bangladesh. The history of this institution dates back to the days of Dhaka Survey School which was established at Nalgola, in Old Dhaka in 1876 to train Surveyors for the then Government of Bengal of British India. As the years passed, the Survey School became the Ahsanullah School of En-

gineering offering three-year diploma courses in Civil, Electrical and Mechanical Engineering. In recognition of the generous financial contribution from the then Nawab of Dhaka, it was named after his father Khawja Ahsanullah. It moved to its present premises in 1912. In 1947, the School was upgraded to Ahsanullah Engineering College as a Faculty of Engineering under the University of Dhaka, offering four-year bachelors courses in Civil, Electrical, Mechanical, Chemical and Metallurgical Engineering. In order to create facilities for postgraduate studies and research, Ahsanullah Engineering College was upgraded to the status of a University in 1962 and was named East Pakistan University of Engineering and Technology. After the War of Liberation in 1971, Bangladesh became an independent state and the university was renamed as the Bangladesh University of Engineering and Technology.

5.4 Students

Till today, it has produced around 25,000 graduates in different branches of engineering and architecture, and has established a good reputation all over the world for the quality of its graduates, many of whom have excelled in their profession in different parts of the globe. It was able to attract students from countries like India, Nepal, Iran, Jordan, Malaysia, Sri Lanka, Pakistan and Palestine.

5.5 Departments

Both Undergraduate and Postgraduate studies and research are now among the primary functions of the University. Eleven departments under five faculties offer Bachelor Degrees, while most of the departments and institutes offer Master's Degrees and some of the departments have Ph.D. programs. In addition to its own research programs, the university undertakes research programs sponsored by outside organizations like European Union, UNO, Commonwealth, UGC, etc. The expertise of the University teachers and the laboratory facilities of the University are also utilized to solve problems and to provide up-to-date engineering and technological knowledge to the various organizations of the country.

Chapter 6

k-safe Labeling of Petersen Graph

In 1898, Petersen produced a trivalent graph with no leaves, now called the Petersen graph [27]. In this chapter we study k-safe labeling for the Petersen graph. We also give upper bound for the span of the Petersen graph. We provide necessary proof for the upper bound.

References

[1] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, "Machine learning: An artificial intelligence approach," *Springer Science and Business Media*, 2013.

[2] M. M. Akbar, M. S. Rahman, M. Kaykobad, E. G. Manning, and G. C. Shoja, "Solving the multidimensional multiple-choice knapsack problem by constructing convex hulls," *Computers & operations research*, vol. 33, no. 5, pp. 1259–1273, 2006.

[3] R. Karim, M. M. Al Aziz, S. Shatabda, M. S. Rahman, M. A. K. Mia, F. Zaman, and S. Rakin, "CoMOGrad and PHOG: From computer vision to fast and accurate protein tertiary structure retrieval," *Scientific reports*, vol. 5, 2015.

[4] M. S. Alam, M. M. Islam, X. Yao, and K. Murase, "Diversity guided evolutionary programming: A novel approach for continuous optimization," *Applied soft computing*, vol. 12, no. 6, pp. 1693–1707, 2012.

[5] M. Kaykobad, "On nonnegative factorization of matrices," *Linear Algebra and its applications*, vol. 96, pp. 27–33, 1987.

[6] M. Kaykobad, "Positive solutions of positive linear systems," *Linear algebra and its applications*, vol. 64, pp. 133–140, 1985.

[7] M. Kaykobad, M. M. Islam, M. E. Amyeen, and M. M. Murshed, "3 is a more promising algorithmic parameter than 2," *Computers & Mathematics with Applications*, vol. 36, no. 6, pp. 19–24, 1998.

[8] "The MCNC benchmark problems for VLSI floorplanning." Last accessed on July 21, 2014, at 02:08:00PM. [Online]. Available: https://www.mcnc.org/.

[9] P.-N. Guo, T. Takahashi, C.-K. Cheng, and T. Yoshimura, "Floorplanning using a tree representation," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 281–289, 2001.

[10] J. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

[11] U. Aickelin and D. Dasgupta, "Artificial immune systems," in *Search Methodologies*, pp. 375–399, Springer, 2005.

[12] J. R. Al-Enezi, M. F. Abbod, and S. A. Sharhan, "Artificial immune systems-models, algorithms and applications," *International Journal*, vol. 3, no. 2, 2010.

[13] S. A. Faruque, M. A. Khatun, and M. S. Rahman, "Modelling direct marketing campaign on social networks," *International Journal of Business Information Systems*, vol. 22, no. 4, pp. 422–435, 2016.

[14] S. Durocher, D. Mondal, and M. S. Rahman, "On graphs that are not PCGs," *Theoretical Computer Science*, vol. 571, pp. 78–87, 2015.

[15] M. S. Rahman, A. Alatabbi, T. Athar, M. Crochemore, and M. S. Rahman, "Absent words and the (dis) similarity analysis of DNA sequences: an experimental study," *BMC research notes*, vol. 9, no. 1, p. 1, 2016.

[16] T. Hashem, L. Kulik, and R. Zhang, "Countering overlapping rectangle privacy attack for moving kNN queries," *Information Systems*, vol. 38, no. 3, pp. 430–453, 2013.

[17] S. M. Farhad, M. A. Nayeem, M. K. Rahman, and M. S. Rahman, "Mapping stream programs onto multicore platforms by local search and genetic algorithm," *Computer Languages, Systems & Structures*, vol. 46, pp. 182–205, 2016.

[18] S. M. B. Malek, M. M. Sadik, and A. K. M. Rahman, "On balanced k-coverage in visual sensor networks," *Journal of Network and Computer Applications*, vol. 72, pp. 72–86, 2016.

[19] G. M. M. Bashir, A. S. M. L. Hoque, and B. C. D. Nath, "E-learning of PHP based on the solutions of real-life problems," *Journal of Computers in Education*, vol. 3, no. 1, pp. 105–129, 2016.

[20] M. Y. S. Uddin and R. Rafiq, "Citizen assisted environmental pollution measurement in developing cities," *International Journal of Environmental Science and Development*, vol. 5, no. 1, p. 70, 2014.

[21] A. Kamal and M. M. Islam, "Boosting up the data hiding rate through multi cycle embedment process," *Journal of Visual Communication and Image Representation*, vol. 40, pp. 574–588, 2016.

[22] M. E. Haque and A. K. M. Rahman, "On constructing interference-aware k-fault resistant topologies for wireless ad hoc networks.," *Ad Hoc & Sensor Wireless Networks*, vol. 19, no. 1-2, pp. 67–94, 2013.

[23] M. S. H. Mukta, M. E. Ali, and J. Mahmud, "Identifying and validating personality traits-based homophilies for an egocentric network," *Social Network Analysis and Mining*, vol. 6, no. 1, p. 74, 2016.

[24] M. E. Ali, E. Tanin, P. Scheuermann, S. Nutanong, and L. Kulik, "Spatial consensus queries in a collaborative environment," *ACM Transactions on Spatial Algorithms and Systems*, vol. 2, no. 1, p. 3, 2016.

[25] M. M. Islam, M. A. Sattar, M. F. Amin, X. Yao, and K. Murase, "A new constructive algorithm for architectural and functional adaptation of artificial neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1590–1605, 2009.

[26] A. A. Al Islam, C. S. Hyder, H. Kabir, M. Naznin, *et al.*, "Stable sensor network (ssn): a dynamic clustering technique for maximizing stability in wireless sensor networks," *Wireless sensor network*, vol. 2, no. 07, p. 538, 2010.

[27] D. A. Holton and J. Sheehan, *The Petersen Graph*, vol. 7. Cambridge University Press, 1993.

Index

1971, *see* War of Liberation

Ahsanullah School of Engineering, 18

BUET, 17
    auditorium, 17
    History, 17
    postgraduate, 17
    undergraduate, 17

Commonwealth, 18

Dhaka, 17

India, 18
Iran, 18

Jordan, 18

Malaysia, 18

Nalgola, 17
Nepal, 18

Pakistan, 18
Palestine, 18

Sri Lanka, 18

UGC, 18

War of Liberation, 18

Appendix A

Algorithms

A.1 Sample Algorithm

In Algorithm 1 we show how to calcute $y = x^n$.

Algorithm 1 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$
Ensure: $y = x^n$
  $y \leftarrow 1$
  if $n < 0$ **then**
    $X \leftarrow 1/x$
    $N \leftarrow -n$
  else
    $X \leftarrow x$
    $N \leftarrow n$
  end if
  while $N \neq 0$ **do**
    if N is even **then**
      $X \leftarrow X \times X$
      $N \leftarrow N/2$
    else $\{N$ is odd$\}$
      $y \leftarrow y \times X$
      $N \leftarrow N - 1$
    end if
  end while

Appendix B

Codes

B.1 Sample Code

We use this code to find out...

```
1 #include <stdio.h>
2 int Fibonacci(int);
3
4 main()
5 {
6   int n, i = 0, c;
7
8   printf("Enter the value of n: ");
9   scanf("%d",&n);
10
11  printf("\nFibonacci series\n");
12
13  for (c = 1 ; c <= n ; c++)
14    {
15      printf("%d\n", Fibonacci(i));
16      i++;
17    }
18
19  return 0;
20 }
21
22 int Fibonacci(int n)
23 {
```

```
24   if (n == 0)
25     return 0;
26   else if (n == 1)
27     return 1;
28   else
29     return (Fibonacci(n-1) + Fibonacci(n-2));
30 }
```

B.2 Another Sample Code

```
1  SELECT associations2.object_id, associations2.term_id,
2         associations2.cat_ID, associations2.term_taxonomy_id
3  FROM (SELECT objects_tags.object_id, objects_tags.term_id,
4       wp_cb_tags2cats.cat_ID, categories.term_taxonomy_id
5  FROM (SELECT wp_term_relationships.object_id,
6       wp_term_taxonomy.term_id, wp_term_taxonomy.term_taxonomy_id
7  FROM wp_term_relationships
8  LEFT JOIN wp_term_taxonomy ON
9       wp_term_relationships.term_taxonomy_id =
10      wp_term_taxonomy.term_taxonomy_id
11 ORDER BY object_id ASC, term_id ASC)
12 AS objects_tags
13 LEFT JOIN wp_cb_tags2cats ON objects_tags.term_id =
14      wp_cb_tags2cats.tag_ID
15 LEFT JOIN (SELECT wp_term_relationships.object_id,
16      wp_term_taxonomy.term_id as cat_ID,
17      wp_term_taxonomy.term_taxonomy_id
18 FROM wp_term_relationships
19 LEFT JOIN wp_term_taxonomy ON
20      wp_term_relationships.term_taxonomy_id =
21      wp_term_taxonomy.term_taxonomy_id
22 WHERE wp_term_taxonomy.taxonomy = 'category'
23 GROUP BY object_id, cat_ID, term_taxonomy_id
24 ORDER BY object_id, cat_ID, term_taxonomy_id)
25 AS categories on wp_cb_tags2cats.cat_ID = categories.term_id
26 WHERE objects_tags.term_id = wp_cb_tags2cats.tag_ID
27 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
28 ORDER BY object_id ASC, term_id ASC, cat_ID ASC)
29 AS associations2
30 LEFT JOIN categories ON associations2.object_id =
```

```
31      categories.object_id
32 WHERE associations2.cat_ID <> categories.cat_ID
33 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
34 ORDER BY object_id, term_id, cat_ID, term_taxonomy_id
```

Generated using Undegraduate Thesis LaTeX Template, Version 1.3. Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Sunday 27th August, 2017 at 11:44am.