# Lecture\_4b

Hello and welcome to the second lecture in this week. We are going to be discussing clustering algorithms. In this lecture video, we are going to look at a recap of unsupervised learning and we are going to examine some common clustering algorithms.

Again, unsupervised learning works mainly by finding intrinsic structure or patterns in data that can be used to draw inferences for datasets that do not have labelled responses. As a rule of thumb, when there is no existing data for how the dataset might be categorised or grouped, then unsupervised learning is to be used. Unsupervised learning is mainly implemented in the form of cluster analysis. Generically, in cluster analysis data is partitioned into clusters (or groups) according to the measure of similarity or shared attributes between the observations in the dataset and clusters are generally formed in such a way that objects in the same cluster are very similar and objects in different clusters are very dissimilar. Clustering analysis usually takes two forms: hard clustering, where each data point in the dataset belongs to only one cluster. Soft clustering, where each data point in the data set can belong to several clusters.

We have plenty of algorithms which can be used to implement hard clustering and soft clustering.

The common hard clustering algorithms include but are not limited to: k-means, k-medoids, hierarchical clustering. self-organising map. The common soft clustering algorithms including but not limited to: fuzzy c-means, gaussian mixture model. K-means: k-means is a form of hard clustering and it mainly works by partitioning the dataset into k number of mutually exclusive clusters, as illustrated. In k-means, how well the data point fits into a cluster is determined by the distance (typically the Euclidean distance) of the point to the cluster's centre, also referred to as the centroid, as illustrated here for cluster one, cluster two and cluster three. Here's a brief overview of how k-means works. The programmer or data analyst decides the number of clusters (k) a priori. Then k random points are selected as the centroids from the data. Then the distance of each data point in the dataset of each centroid is computed and the nearest or closest cluster centroid to each data point is assigned to that data point. Then the same centroid of newly formed clusters are calculated again and these are assigned as the new centroids. The third and forth steps are repeated until the centroids of newly found clusters do not change or the maximum number of iterations are reached. K-means is best used when the number of clusters is known and for fast clustering of very large datasets.

There isn't a proven optimal way of deciding the value of k, which the programmer or analyst has to specify a priori. At best, analysts and scientists simply make an educated guess. K-medoids. This is also a form of hard clustering and it's very similar to k-means. The number of clusters (k) is also specified a priori by the programmer or data analyst. K-medoids makes use of actual data points in the dataset as centres, also called medoids or exemplars as illustrated here. As a result, there is greater interpretability of the cluster centres than in k-means. In k-means, the centre of the cluster is not necessarily one of the input data points. It's always the average between the points in the cluster as illustrated here. To implement k-medoids, a sum of pairwise dissimilarities is minimised and this makes it more robust to noise and outliers in comparison to k-means which minimises a sum of squared Euclidean distances. K-medoids is best used when the number of clusters is known and to scale to large datasets. Also for fast clustering of categorical data. Hierarchical clustering. Hierarchical clustering premises on k-means clustering. However, it takes away the problem of having to pre-define or specify the number of clusters a priori. Hierarchical clustering work by building nested sets of clusters and it analyses the similarities between pairs of data points and groups objects into a binary hierarchical tree. Simply put, it combines the most similar clusters together and repeats the process until only a single cluster is left. Hierarchical clustering can be performed in two main ways.

Agglomerative hierarchical clustering and divisive hierarchical clustering. For divisive hierarchical clustering, if there are n data points in the dataset, all the data points are assigned to a single cluster

to have just one cluster initially, and at each iteration, the farthest point in the cluster is split or divided. The process is repeated until each cluster only contains a single data point. In other words, at the end we have n clusters. As illustrated here for five data points assigned to a single cluster initially and at the end we have single cluster each.

For agglomerated hierarchical clustering, which is the most widely used clustering method, if there are any data points each data point is assigned to a cluster to have an individual cluster to have n clusters in total initially. Then the smallest distance in the proximity matrix, which is an n-by-n matrix formed where all the elements are the respective distances of the data points from each other is deduced. The points with a smaller distance are merged and the proximity metrics is updated. The process is repeated until there is only one cluster left and as you can see, agglomerated hierarchic clustering is just the reverse process of divisive clustering. Hierarchical clustering is used when the number of clusters in the dataset is not known beforehand, and it's also used to have a visualisation to guide the selection using what we call a dendrogram. Self-organising map is a hard clustering algorithm. It is a neural network-based clustering method which works by transforming a dataset into a topology-preserving low-dimensional, typically two dimensional discretised representation of the dataset, that is input space of the training samples and this representation is called a map. From this definition, we can tell that self-organising map is a dimensionality reduction method.

Self-organising maps differ from other artificial neural networks due to the adoption of competitive learning as opposed to error correction learning such as backpropagation with gradient descent,

which is found in conventional artificial neural networks. In competitive learning, the nodes in the neural network compete for the right to respond to a subset of the input data. Self-organising map is best used to: visualise high dimensional data in 2D or 3D. They are so used to reduce the dimensionality of data by preserving its topology or shape of the data.

Fuzzy c-means. Fuzzy c-means is a soft clustering algorithm and it is a partition-based form of clustering that allows data points from the dataset to belong to more than one cluster. It is very similar to k-means algorithm, however it attempts to partition a finite collection of data points into a collection of c fuzzy clusters according to some given criterion. Fuzzy c-means also makes use of membership values and a fuzzifier which determines the level of cluster fuzziness. Intuitively, we can tell that fuzzy c-means premises on fuzzy logic. Fuzzy c-means is best used when the number of clusters is known and the clusters overlap. Fuzzy c-means is also used for pattern recognition.

Gaussian mixture model. Gaussian mixture model is a soft clustering algorithm that allows data points from different multivariate normal distributions with some certain probabilities to belong to several clusters. It is also a partition-based clustering method which is very similar to model-based density estimation. In model-based density estimation, just one multivariate normal distribution, that is gaussian distribution is used to model the probability density function of the unknown probability distribution. The dataset we are working with must have been drawn from. In gaussian mixture model, a weighted sum of several multivariate normal distributions is used instead. Gaussian mixture model is best used when a data point in a dataset may belong to more than one cluster and when we have clusters with different sizes and correlations. In this video, we've looked at a recap of unsupervised learning and we've looked at common cluster algorithms like k-means, k medoids, hierarchical clustering, fuzzy c-means and gaussian mixture model.