IAML DL - Study Guide - Week 08

Sambit Paul, Pavlos Andreadis

March 2020

1 Introduction

Week 8 extends on clustering from Week 7 and speaks on how to create multi-level hierarchy of clusters and introduces one of the most simple dimensionality reduction techniques - Principal Component Analysis.
Principal Component Analysis (PCA) is an unsupervised method to reduce the number of dimensions/features such that the reduced number of dimensions still contain most of the information from the large set.
Hierarchical Clustering is an extension to the basic clustering techniques discussed in Week 7. Hierarchical clustering uses two methods of clustering - joining nodes to form multi-level clusters, and splitting big clusters into smaller units. These methods are called agglomerative and divisive clustering respectively.

2 Principal Component Analysis

- A primer for dimensionality reduction techniques and an introduction to various techniques can be found in this article.
- One of the key concepts behind the need of PCA originates from the idea of **The Curse of Dimensionality**. Bishop [2006] explains the concept in Section 1.4 and this video playlist provides a very intuitive guide to the problems relating to the curse of dimensionality.
- Details of the mathematics behind PCA can be found in Section 15.2 of Barber [2012].
- Jolliffe and Cadima [2016] is a very good paper to read about PCA and understand it using examples. Section 2a of this paper gives key insights into how PCA is used for data analysis. This is supported by the example provided in Section 2b.
- A simpler explanation of PCA with code [Matlab & Python] is provided in Section 2 of this article. This article also talks about projection in other dimensions.

- Often while reading about PCA, the term eigen decomposition will come up. To understand the role eigen values and eigen decomposition plays in PCA, please use this article.
- Singular Value Decomposition is a method to perform PCA and this article talks about the fundamentals of SVD. This explanation here is a very good mathematical explanation of how SVD and PCA are related.
- Linear Discriminant Analysis (LDA) is a supervised method of dimensionality reduction which uses class information and a Gaussian distribution assumption to project to a lower dimension. Please use this article to know more about Fisher's Linear Discriminants.
- A very succinct introduction of LDA is provided in this article. To read about it in depth, please use Barber [2012] Section 16.2.
- A comparison between PCA and LDA is provided in this article.

3 Hierarchical Clustering

- **Need** of Hierarchical Clustering:
  - No assumptions need to be made about the number of clusters.
  - The generated dendograms tend to correspond to meaningful taxonomies.
- **Types** of Hierarchical Clustering:
  - *Agglomerative Hierarchical Clustering*: Initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed.
  - *Divisive Hierarchical Clustering*: All the data points are considered a single cluster and in each iteration, data points which are not similar are separated from the cluster. Each data point which is separated is considered as an individual cluster.
- To have a deeper understanding of agglomerative clustering, please refer to Patel et al. [2015] which gives an introduction to low complexity methods like **CURE**, **BIRCH**, and linkage algorithms like **SLINK**, **AVELINK** and **CLINK**.
- The algorithm for divisive clustering is as follows Roux [2018]:
  1. Splitting procedure for the subdivision of clusters into two sub-clusters
  2. Local evaluation of the bipartitions resulting from the tentative splits
  3. Formula for determining the node levels of the resulting dendrogram

- Roux [2018] provides a very good explanation of divisive clustering explaining each of the steps of the algorithm and exploring the different methods employed at each step.

- Section 2.1 of this article gives a very good intuition of **Ward's method**.

References

David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

Christopher M Bishop. *Pattern recognition and machine learning*. Springer Science+ Business Media, 2006.

Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

Sakshi Patel, Shivani Sihmar, and Aman Jatain. A study of hierarchical clustering algorithms. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 537–541. IEEE, 2015.

Maurice Roux. A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, 35(2):345–366, 2018.