

# Unsupervised Learning-Summary Sheet

ML@LSE

2018/2019

## Abstract

**Objectives:** Understand two important unsupervised learning methods: Principal Component Analysis and Clustering. Understand the difference between k-means clustering and Hierarchical Clustering.

**Requirements:** Introductory Bootcamp (you can read the slides if you didn't attend). Although not necessary, familiarity with notions of linear algebra may help: inner product and orthogonal projections. Strong understanding of the summation operator ( $\sum_{i=1}^n, \sum_{j \in C}$ ) may help.

**Keywords:** PCA, Loading vector, Principal Components, Proportion of variance explained, Clustering, k-means Clustering, Within-cluster Variation, Hierarchical Clustering, Minimal Intercluster Dissimilarity, Dendrogram.

## *A PRINCIPAL COMPONENT ANALYSIS*

---

### **A.1 Intuition behind PCA**

- a How to represent a Galaxy in 2D?
- b PCA main idea: maximize the variance

### **A.2 PCA Algorithm**

- a Finding the principal components
- b Proportion of variance explained

## *B K-MEANS CLUSTERING*

---

### **B.1 The intuition behind K-MC**

- a Why clustering?
- b K-MC partitions our data

### **B.2 K-MC Algorithm**

- a Within-Cluster variation
- b K-MC in practice

## *C HIERARCHICAL CLUSTERING*

---

### **C.1 The intuition behind HC**

- a A more flexible approach
- b Agglomerative Clustering

### **C.2 HC Algorithm**

- a Summary of the Algorithm
- b Dendrograms