KLIC

Stephen Coleman

25/11/2019

Kernel learning integrative clustering

Elevator pitch

Takes independent clusterings of the individuals (i.e. local clusterings) and combines these for a global clustering. Allows different local clusterings to contribute with different strenghts to the final clustering. No specifications on the type of model used to create the original clusterings.

Intro

Kernel Learning Integrative Clustering (KLIC) is a **sequential analysis** method of integrative clustering. This in comparison to **post-processing** or **joint** methods. Are these labels a real thing?

KLIC is an extension of Cluster-of-Cluster Analysis (COCA). If one has L different datasets of measurements for the same N individuals to which one applies independent clustering methods, COCA then turns the similarity matrices that result from this into a global clustering by combining the matrices in a method similar to Consensus Clustering (the original paper). KLIC extends this by allowing different similarity matrices to have different weights in how they contribute to the global clustering. In short, KLIC applies multiple kernel k-means clustering to similarity matrices generated for individual datasets.

To understand KLIC one must understand the following:

- COCA;
- the kernel trick;
- k-means clustering; and
- multiple kernel k-means clustering.

The kernel trick

This is a computational trick to avoid operations. It aims to do an analysis in a high-dimensional space while only considering calculations in the original space.

Definition: Positive definite kernel

or simply a kernel, δ, is a symmetric map:

$$\delta : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

which for all $x_1, \ldots, x_N \in \mathcal{X}$, the matrix Δ defined by entries $\Delta_{ij} = \delta(x_i, x_j)$, is positive semi-definite.

Definition: Kernel matrix

or **Gram matrix**, Δ, is the positive semi-definite matrix defined by a kernel δ applied to data $\mathcal{X} = (x_1, \ldots, x_N)$ with entries $\Delta_{ij} = \delta(x_i, x_j)$.

Definition: Feature map

For each kernel δ there exists a **feature map** $\phi(\cdot)$, which maps the original data $\mathcal{X} = (x_1, \ldots, x_N)$ to some new feature space taking values in some inner product of $\mathcal{X}$ defined by:

$$\delta(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle.$$

Thus if one is interested in working in some feature space that can be defined in terms of inner products, one may use kernels to avoid computations in analysing the data in said space.

k-means clustering

Imagine data and cluster it using some k centroids. Hooray.