Convex Clustering Example

D2K Course Staff

February 16 2019

Introduction

One of them newfangled clustering methods that all the cool stat ids are using these days is called convex clustering. The basic idea is to minimize the sum of the distances between each of the observations and their assigned cluster centroid and the distances between each of the assigned cluster centroids. This can be thought of as something like a ridge or LASSO problem for clustering.

Formally, define our data matrix as $\mathbf{X}$ with $i \in 1, ..., m$ rows of observations and $j \in 1, ..., n$ columns. We then want to find the corresponding cluster centroid matrix $\mathbf{U}$, also with $i \in 1, ..., m$ rows and $j \in 1, ..., n$ columns, that minimizes

$$\frac{1}{2}\sum_{i=1}^{m} \|\mathbf{X}_i - \mathbf{U}_i\|_2 + \gamma \sum_{k<l} w_{kl} \|\mathbf{U}_k - \mathbf{U}_l\|_2.$$

In the above, $\mathbf{U}_i$ is the centroid that corresponds to $\mathbf{X}_i$ (not necessarily unique) and $\|.\|_2$ is your typical Euclidean distance. w_{kl} is the weight that we give to each centroid pairing distance term - in general, this is calculated such that pairs of observations of $\mathbf{X}_i$ that are further away have smaller weights (intuitively, we care less about fusing cluster centroids for points that are far apart compared to points that are close to each other.) γ is the regularization parameter than controls the balance of the trade-off between the two summation terms. In the same way that the LASSO regulaization parameter pulls each coefficient toward 0, the convex clustering regularization parameter fuses the cluster centroids together toward the same point.

This has a few advantages over your basic clustering techniques:

1. The optimization problem has a unique solution for each given value of γ and can be implemented via known convex opimization algorithms. This also allows us to choose the number of cluster centroids in a data-driven manner by setting γ such that the solution gives the desired number of unique centroids.
2. We get an intuitive solution to changing the number of cluster centroids - decreasing the number of clusters fuses separate clusters together and increasing the number of clusters splits clusters apart. (Compare thta to k-means in the basic clustering example.)
3. We can create a regularization path, akin to ridge or LASSO regression to see how clusters fuse or split apart as we decrease or increase the number of clusters.

Implementation

There are a couple of R package implementations for convex clustering, `cvxclustr` and `cclust`. See which one works best for you. Read the CRAN reference for these for more details.

No Examples?

Sorry. We're trying to find a good, simple example that gives an intuitive answer. For now, use the examples in the corresponding R package references.