Medoid in squared ℓ_2 norm

Yao Xie

August 2020

Consider the following scenario. There are m data points: $x^1, \ldots, x^m$, all n-dimensional vectors. We will prove that under squared-ℓ_2 norm, the medoid correspond to the data point that is closest to the mean. Thus, *in this special case*, finding solution to k-medoid, we can perform average of the data point in the cluster (as in k-means) and then finding the data point closest to it if using squared ℓ_2 norm. Please note that this conclusion may NOT generalize to other dissimilarity metrics (even may not generalize to ℓ_2 norm without the square).

Consider the following problem of finding a centroid the data point that minimizes the sum-of-squared ℓ_2 distance

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{m} \|x^i - \mu\|^2$$

As can be shown, the minimizer is given by the mean

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} x^i.$$

Now consider the following problem of finding medoid in the squared ℓ_2 norm (finding a data point that is closest to everyone else in the squared ℓ_2 norm):

$$M = \arg \min_{\{x^j, j=1,\ldots,m\}} \sum_{i=1}^{m} \|x^j - x^i\|^2$$

Note that the cost function can be written as

$$\begin{aligned}
&\sum_{i=1}^{m} \|x^j - x^i\|^2 \\
&= m\|x^j\|^2 + \left(\sum_{i=1}^{m} \|x^i\|^2\right) - 2\left(\sum_{i=1}^{m} (x^i)^T x^j\right) \qquad (1) \\
&= m\|x^j\|^2 + \left(\sum_{i=1}^{m} \|x^i\|^2\right) - 2m\mu^T x^j
\end{aligned}$$

Note that we can vary x^j, the second term does not matter, so we can drop it from consideration. After dropping constant m which does not affect the optimal solution, this means

$$M = \arg \min_{\{x^j, j=1,\ldots,m\}} \|x^j\|^2 - 2\mu^T x^j$$

On the other hand, consider the problem of finding the closest point to the mean in squared ℓ_2 dissimilarity metric (for a given $\hat{\mu}$):

$$M' = \arg \min_{\{x^j, j=1,\ldots,m\}} \|x^j - \hat{\mu}\|^2$$

Using similar expansion of the squared ℓ_2 norm:

$$\|x^j - \hat{\mu}\|^2 = \|x^j\|^2 + \|\hat{\mu}\|^2 - 2\hat{\mu}^T x_j$$

and dropping a term $\|\hat{\mu}\|^2$ which does not affect the solution we have

$$M' = \arg \min_{\{x^j, j=1,\ldots,m\}} \|x^j\|^2 - 2\hat{\mu}^T x^j.$$

This has shown $M = M'$, and thus finding the medoid in **squared** ℓ_2 distance is equivalent to finding the mean of the data point and then finding the closest data point to it.