Affinity Propagation

Definition and purpose

**Affinity propagation (AP)** is a clustering algorithm based on the concept of "**message passing**" between data points**. It does not require the number of clusters to be determined or estimated before running the algorithm (differ from the prototype-based clustering algorithms).**

At first, we need to define the similarity between node and , and it should satisfy that if and only if is more similar to than to . So, we use the negative squared distance of two data points, which as

**The diagonal of (i.e., )** is a hyper-parameter which represents the **instance preference**, meaning how likely a particular instance is to become an exemplar(prototype). **When it is set to the same value for all inputs, it controls how many classes the algorithm produces.** Smaller value produces fewer classes, vice versa. We usually use the median.

The Process is

1. **Each node, initially, is a cluster**
2. At each round, **we merge any cluster to the cluster that is closest to it**
3. We stop the algorithm when we have the desired number of clusters

The algorithm proceeds by alternating between two message-passing steps, which updates two matrices:

1. The “Responsibility” “sent” from to
2. The “Availability” “sent” from candidate to point

Examples

if we have the following dataTable

Description automatically generated

Treat every node as one cluster, and calculate the similarity matrix using

Table

Description automatically generated