

Zad 1.

środa, 14 czerwca 2023 11:06

Zadanie 1. Na wykładzie mówiliśmy o algorytmie K -średnich (K -means). Istnieją jeszcze dwa inne algorytmy o podobnych nazwach: K -medians, K -medoids. Opisz je, koncentrując się na różnicach z K -means¹.

¹Zadanie wymaga samodzielnego poszukania definicji algorytmów, zachęcam jednak, aby spróbować samemu wymyślić, co te nazwy oznaczają, przed rozpoczęciem poszukiwań

K -średnich

Przez cały czas działania algorytmu pamiętamy K **prototypów** (czyli punktów będących reprezentantami grupy)

Algorytm przeplata dwie fazy:

1. Przypisanie każdego punktu do najbliższego mu prototypu
2. Obliczenie nowych prototypów jako **średnich** wszystkich punktów przypisanych do tego samego prototypu

- Miara: błąd średniokwadratowy, czyli

$$\sum_{x \in \text{Dane}} (x - \text{reprezentant}(x))^2$$

1) k -median różnica - motywacja

In statistics, **k -medians clustering**^{[1][2]} is a **cluster analysis** algorithm. It is a variation of **k -means clustering** where instead of calculating the **mean** for each cluster to determine its centroid, **one instead calculates the median**. This has the effect of minimizing error over all clusters with respect to the **1-norm distance metric**, as opposed to the squared **2-norm distance metric** (which **k -means** does.)

↑
manhattan

2) k -medoids

In contrast to the k -means algorithm, **k -medoids** chooses **actual data points as centers (medoids or exemplars)**, and thereby allows for greater interpretability of the cluster centers than in **k -means**, where the **center of a cluster is not necessarily one of the input data points** (it is the average between the points in the cluster). Furthermore, **k -medoids can be used with arbitrary dissimilarity measures**, whereas **k -means generally requires Euclidean distance** for efficient solutions. Because **k -medoids minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances**, **it is more robust to noise and outliers than k -means**.

... A medoid can be defined as a point in the cluster, whose dissimilarities with all the other points in the cluster are minimum.

Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points.

