

Inteligență computațională

Tema

Algoritmul *K-means clustering* este o metodă de determinare a clusterelor pe care le formează mai multe pattern-uri. Procedura este una de instruire nesupervizată. Se presupune cunoscut numărul K al clusterelor, acesta fiind un parametru stabilit *a priori*.

Fiecare cluster are un centroid. Algoritmul lucrează cu K cluster, deci K dintre punctele folosite la instruire vor fi centriozii celor K cluster. Întrucât inițializarea centrozilor se face aleator, există posibilitatea ca mai multe rulări ale algoritmului să conducă la rezultate diferite.

Fiecare punct este asociat clusterului determinat de cel mai apropiat centroid. Distanța dintre punct și centroid poate fi calculată, de exemplu, ca distanță euclidiană, dar se poate opta și pentru alte variante.

Algoritmul este următorul:

1. *Se alege aleator K puncte ca centroizi inițiali.*
2. *Se formează K cluster prin asignarea tuturor punctelor celor mai apropiați centroizi.*
3. *Se recalculează centrozii astfel: noul centroid va fi centrul de greutate determinat de punctele clusterului.*
4. *Se reiau pașii 2 și 3 până când centrozii nu se mai modifică.*

Aplicați algoritmul *K-means clustering* pentru următoarele puncte, cu $K=3$:

P1: 45 85
P2: 50 43
P3: 40 80
P4: 55 42
P5: 200 43
P6: 48 40
P7: 195 41
P8: 43 87
P9: 190 40

Listați cei 3 centroizi și punctele asociate fiecărui cluster.

Pentru o descriere detaliată a algoritmului puteți consulta și:

<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxzZWpveXNhaGFzaXRlfGd4OjJhMjI5ZTU1ZTgxMDRhYjk>