

K-Means Klusterointi

- Yksinkertainen ja nopea algoritmi
- Soveltuu numeeriselle datalle

Vaiheet

1. Valitaan K (klustereiden lukumäärä)
2. Valitaan sattumanvaraisesti k -kappaletta mean (C) pisteitä
3. Jaetaan pisteet euklidinen etäisyyden avulla. piste kuuluu lähimpään C :n klusteriin
4. Otetaan klustereiden keskiarvo-piste ja tehdään uusi jako
5. Jos C ja klusterit eivät muutu lopetetaan
6. Muutoin palataan kohtaan 4

Edut

- Nopea
- Yksinkertainen toteuttaa
- Vaativuus helposti arvioitavissa

Haitat

- Toimii vain numeeriselle datalle
- Herkkä kohinalle ja outlander-pisteille