

K Nearest Neighbors(KNN)

- Yksinkertainen classifier-algoritmi
- Perustuu datapisteiden etäisyyksien laskemiseen
- Euklidinen etäisyys(Euclidean distance)
- K-arvolla määritellä kuinka monta naapuria otetaan huomioon

- Äänestetään enemmistö voittaa
- K kannattaa yleensä olla pariton luku -> äänet eivät mene tasan
- Jos classeja on paljon K:n valinta vaikeutuu

Edut

- Yksinkertaisuus
- Nopea kouluttaa
- Data on vain dataa-> Ei tee mitään olettamuksia eikä tarvitse ennakkotietoja

Haitat

- Ei oikeasti opi mitään
- Herkkä puuttuvalle datalle, outline pisteille ja “turhalle” datalle
- Hitaus suuren datan kanssa
- K:n valinnan vaikeus
- Tarvitsee paljon tallennustilaa

Mihin käytetään

- Data/text mining
- Ilmastomallit
- Talous/pörssi
- Lääketiede