

# TEK5020/9020 - Mønstergjenkjenning Høsten 2023

Løsningsforslag – Øvingsoppgaver 6

Idar Dyrdal (idar.dyrdal@its.uio.no)

UiO : Institutt for teknologisystemer

23. august 2023

## Oppgave 1a

Den grunnleggende *Isodata-algoritmen* (K-means-clustering) består av følgende trinn:

- Initialiser  $\hat{\mu}_1, \dots, \hat{\mu}_c$
- Gjenta inntil ferdig:
  - Klassifiser  $\mathbf{x}_k, k = 1, \dots, n$  til nærmeste middel
  - Oppdater  $\hat{\mu}_1, \dots, \hat{\mu}_c$
  - Hvis ingen endring  $\Rightarrow$  ferdig.

## Oppgave 1b

Det endimensjonale datasettet gitt ved

$$\mathcal{C} = \{1.5, 1.7, 2.0, 2.1, 2.9, 3.1, 3.8, 4.0\}$$

skal deles i to klynger  $\mathcal{C}_1$  og  $\mathcal{C}_2$  ved bruk av Isodata-algoritmen.

Initialiserer klyngesentrene:

$$\left. \begin{array}{l} \hat{\mu}_1 = 2.0 \\ \hat{\mu}_2 = 4.0 \end{array} \right\} \text{terskel} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} = 3.0$$

## Oppgave 1 (forts.)

Iterasjon 1:

$$\left. \begin{aligned}\hat{\mu}_1 &= \frac{1.5 + 1.7 + 2.0 + 2.1 + 2.9}{5} = 2.04 \\ \hat{\mu}_2 &= \frac{3.1 + 3.8 + 4.0}{3} = 3.63\end{aligned} \right\} \text{terskel} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} = 2.84$$

Iterasjon 2:

$$\left. \begin{aligned}\hat{\mu}_1 &= \frac{1.5 + 1.7 + 2.0 + 2.1}{4} = 1.83 \\ \hat{\mu}_2 &= \frac{2.9 + 3.1 + 3.8 + 4.0}{4} = 3.45\end{aligned} \right\} \text{terskel} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} = 2.64$$

## Oppgave 1 (forts.)

Iterasjon 3:

$$\left. \begin{aligned} \hat{\mu}_1 &= \frac{1.5 + 1.7 + 2.0 + 2.1}{4} = 1.83 \\ \hat{\mu}_2 &= \frac{2.9 + 3.1 + 3.8 + 4.0}{4} = 3.45 \end{aligned} \right\} \text{terskel} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} = \underline{\underline{2.64}}$$

Ingen endring i siste iterasjon  $\Rightarrow$  ferdig.

Den endelige terskelen (desisjongrensen) gir klyngeinndelingen:

$$\mathcal{C}_1 = \{1.5, 1.7, 2.0, 2.1\},$$

$$\mathcal{C}_2 = \{2.9, 3.1, 3.8, 4.0\}.$$

## Oppgave 2

Det univariate datasettet

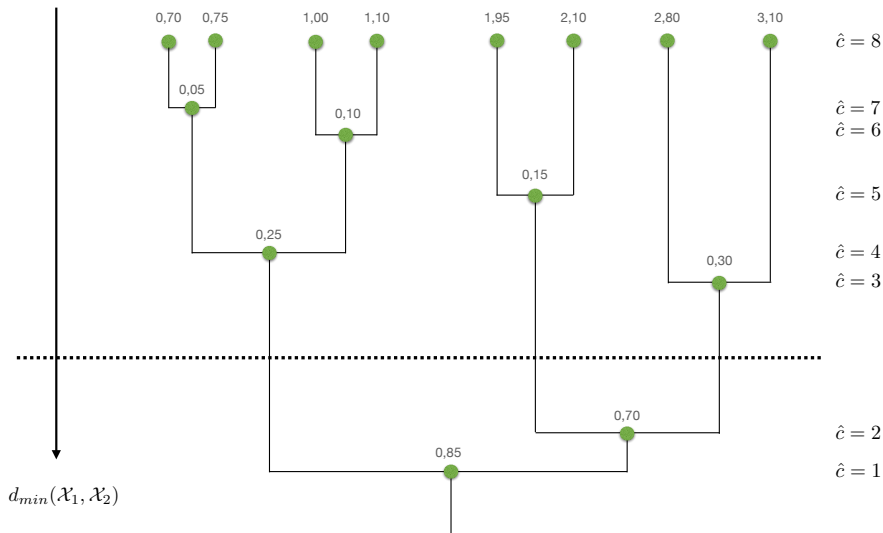
$$\mathcal{X} = \{0.70, 0.75, 1.00, 1.10, 1.95, 2.10, 2.80, 3.10\}$$

skal deles i tre klynger ved å bruke den agglomerative metoden med avstandsmålet  $d_{\min}(\mathcal{X}_1, \mathcal{X}_2)$ . Metoden består av følgende trinn:

- 1 Start med  $n$  klynger, dvs.  $\mathcal{X}_i = \{\mathbf{x}_i\}, i = 1, \dots, c; \hat{c} = n$
- 2 Finn nærmeste par av klynger, f.eks.  $\mathcal{X}_i, \mathcal{X}_j$
- 3 Slå sammen disse;  $\hat{c} \rightarrow \hat{c} - 1$
- 4 Gjenta trinn 2 og 3 inntil ønsket antall klynger  $\hat{c} = c$  er funnet.

Dendrogrammet (neste slide) illustrerer gangen i prosessen i dette tilfellet.

## Oppgave 2 (forts.)



## Oppgave 2 (forts.)

Inndeling i tre klynger (se stiplet linje i dendrogrammet)  
gir da klyngeinndelingen:

$$\mathcal{X}_1 = \{0.70, 0.75, 1.00, 1.10\},$$

$$\mathcal{X}_2 = \{1.95, 2.10\},$$

$$\mathcal{X}_3 = \{2.80, 3.10\}.$$