

Øvingsoppgaver

TEK5020/9020 – Mønstergjenkjenning

Del 6 – Klyngeanalyse

Høsten 2023

Oppgave 1

- a) Skisser den grunnleggende *Isodata-algoritmen* (K-means-clustering).
- b) La datasettet i et klyngeanalyseproblem være mengden av endimensjonale sampler gitt ved

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

Bruk Isodata-algoritmen til å dele dette datasettet i *to* klynger \mathcal{C}_1 og \mathcal{C}_2 . Sett startverdiene for klyngesentrene til $\hat{\mu}_1 = 2.0$ og $\hat{\mu}_2 = 4.0$.

Oppgave 2

La datasettet i et klyngeanalyseproblem være mengden av endimensjonale sampler gitt ved

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

Bruk den agglomerative metoden til å dele \mathcal{X} i *tre* klynger. Bruk avstandsmålet $d_{\min}(\mathcal{X}_1, \mathcal{X}_2)$, dvs. minste Euclidske avstand mellom to sampler fra hver sin klynge \mathcal{X}_1 og \mathcal{X}_2 . Illustrer løsningen ved hjelp av et dendrogram.