Institut for Matematik og Datalogi
Syddansk Universitet, Odense

30. oktober 2017
RF

Studiegruppetimer DM534

Til studiegruppetimerne (TE-SF) behøver man ikke forberede sig, men skal blot mødes og arbejde i gruppen. Dette er anderledes end til eksaminatorietimer (TE), hvor det forventes at man har forsøgt at løse opgaverne inden timerne.

Forslag til SF-timer i uge 44

1. Repetér for hinanden definitionen af en centroide for en cluster. Hvad er centroiden for en cluster C bestående af følgende tre punkter $C = \{(2,3),(5,5),(4,1)\}$? Check beregningen af de to centroider i figuren på side 19 i Richard Röttgers slides.

2. Repetér forskellen på Forgy-Lloyd og MacQueen udgaverne af k-means algoritmen. Giver de to udgaver altid samme resultat?

3. Den velkendte længde (norm) i $\mathbb{R}^2$ (planen) for et punkt $\vec{v} = (x,y)$ er givet ved $\sqrt{x^2+y^2} = (x^2+y^2)^{1/2}$. Dette kaldes også L_2-normen af $\vec{v}$. Mere generelt er L_P-normen af et punkt $\vec{v} = (x_1, x_2, \ldots, x_k)$ i $\mathbb{R}^k$ givet ved $\sqrt[P]{\sum_{i=1}^{k} |x_i|^P} = (\sum_{i=1}^{k} |x_i|^P)^{1/P}$. En norm kan bruges til at definere afstanden mellem to punkter $\vec{p}$ og $\vec{q}$ som længden (normen) af $\vec{p} - \vec{q}$. Dette kaldes så L_P-afstanden.

   Forklar figuren midt på side 73 i Richard Röttgers slides (som har $P = 1$ og $k = 2$). Dvs. forklar hvorfor mængden af alle punkter $\vec{q}$ i en given afstand r fra et punkt $\vec{q}$ (også kaldet kuglen om $\vec{q}$ med radius r) har en sådan facon.

   Man definerer L_∞-normen af $\vec{v} = (x_1, x_2, \ldots, x_k)$ som $\max_{i=1}^{k} |x_i|$. Forklar figuren til højre på side 73 i Richard Röttgers slides. Forklar hvorfor L_∞ er et godt navn, når man sammenligner med definitionen af L_P (hint: for $\vec{v} = (27,2)$ beregn $(27^3 + 2^3)^{1/3}$ og $(27^{10} + 2^{10})^{1/10}$).

   For at få en ide om udseendet af kugler i L_P-normer generelt, se den engelske Wikipedia-side om superellipsen.