

Inlämningsuppgift 5 - Klustering
Maskininlärning och Deep learning - VT 2023
Individuell inlämning
Lämna in senast 24/5 08:00

I denna inlämning ska en egen K-Means skapas.

Inlämningen är uppdelad i en godkänt del och en väl godkänt del.

För godkänt krävs godkänt på godkänt delen, inklusive efter komplettering.

För väl godkänt krävs att man blir godkänd på både godkänt delen och väl godkänt delen vid första inlämningen.

En notebook ska lämnas in som innehåller kod samt svaret på frågor. Tänk på att förklara bilder och svar. Döp filen till *för_efternamn_kmeans.ipynb*

För både godkänt och väl godkänt skall en egen K-Means skrivas. För godkänt skall algoritmen kunna hantera ett dataset och för väl godkänt ska den kunna hantera två olika dataset. För båda nivåerna ska det gå att välja olika värden på K och distansen ska kunna beräknas med euklidisk distans. Det är godkänt att göra det möjligt att välja mellan olika distanser.

För båda nivåerna skall följande punkter och frågor behandlas.

- Skriv en egen K-Means
- För varje dataset
 - Analysera datan
 - Innehåller datan konstiga värden?
 - Ska några attribut plockas bort?
 - Behövs datan normaliseras?
 - Träna din K-Means (experimentera med K)
 - Rita ut klustrena
 - Det är godkänt att rita ut i både 2 dimensioner och 3 dimensioner
 - Använd bara PCA för att rita ut klustrena

Dataset för godkänt:

small_iris.csv

Dataset för väl godkänt:

iris.csv

Mall_Customers.csv