

732G12 Data Mining

Föreläsning 7

Johan Alenlöv

IDA, Linköping University, Sweden

- Introduktion
- K-means klustring
- Hierarkisk klustring

- Ööversakad inlärning: Lära sig data utan responsvariabel!
- Flera olika algoritmer:
 - **Klusteranalys**
 - Associationsanalys
 - Sekventiella mönster
 - Dimensionality reduction techniques
 - PCS, Faktormodeller
 - Representation Learning

Måler med klusteranalys är att dela upp datamaterialet i grupper (kluster) som är intressanta och/eller användbara.

Vi vet inte i förväg vilka grupper som kommer att bildas.

Ingen responsvariabel.

Ta 5 minuter att fundera på följande frågor:

1. Hur många kluster finns det i bilden?



2. Kom på något område där klusteranalys kan vara användbart.

Klusteranalys

Ett "kluster" är inte entydigt definierat.



Tillämpningsområden:

- Biologi (toxonomi/gener)
- Informationssökning (Sökmotorer)
- Psykologi och medicin
- Kunddata
- Sociala medier/nätverk