

Introduktion till Maskininläring

Vad är
Maskininläarning?

Vad är Maskininlärning?

- Maskininlärning = datorn lär sig mönster från data
- Syftet: förutsäga något på ny data
- Vi programmerar inga regler – vi tränar en modell

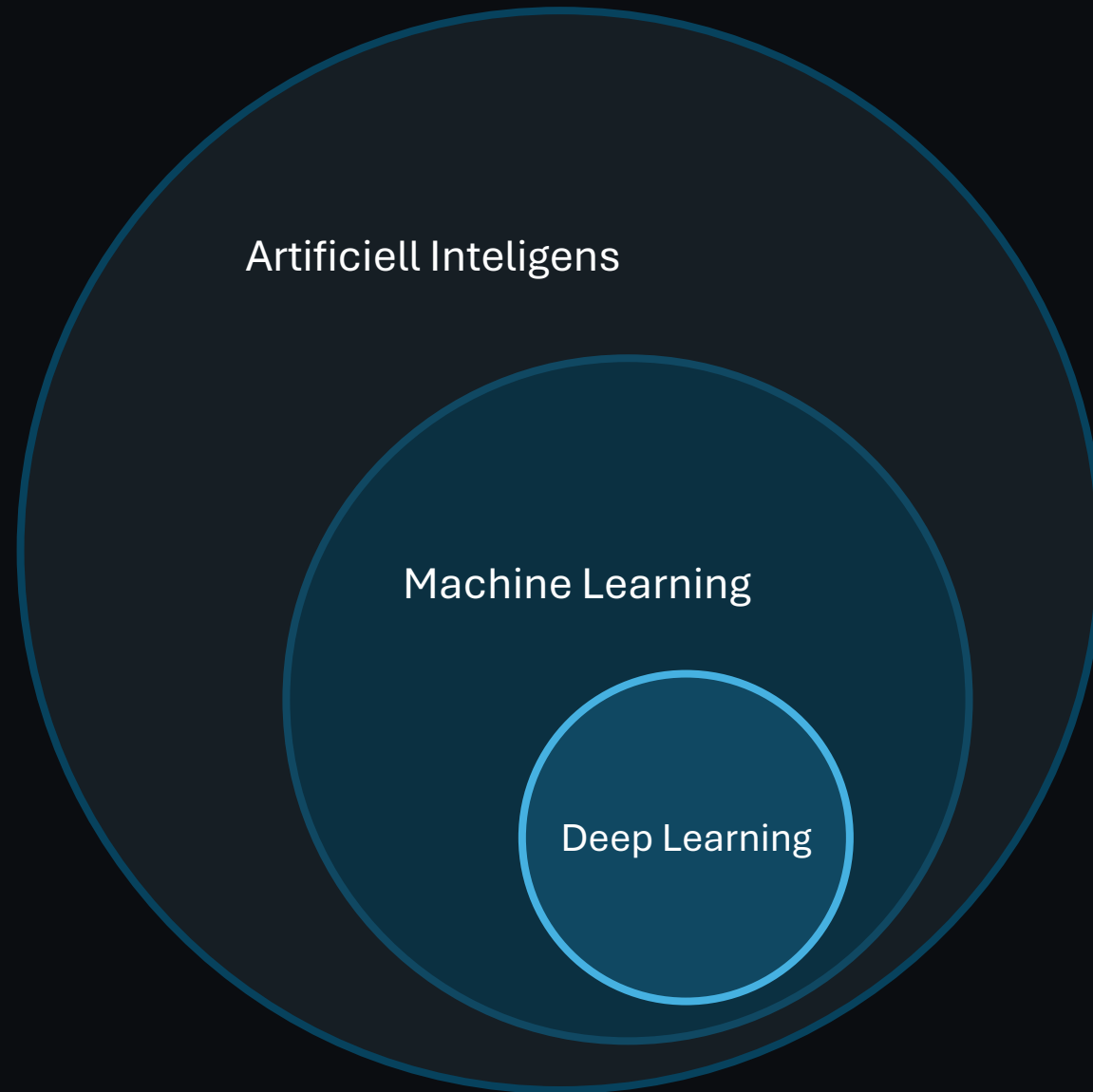
En modell som ”funktion”

- Vi har input (X) \rightarrow modell \rightarrow output (\hat{y})
- X = features (förklarande variabler)
- Y = label/target (det vi vill förutsäga)

Varför inte ”explicit programmerad”?

- Vi väljer en modelltyp
- Vi definierar ett felmått (loss)
- Algoritmen justerar modellen för att minska felet

AI, ML, Deep Learning



Två huvudkategorier

1. Supervised Learning (väglett)
 - Vi har facit (y)
 - Regressionsproblem
 - Klassificeringsproblem
2. Unsupervised Learning (icke-väglett)
 - Vi har inget facit
 - Klustering
 - Dimensionsreducering

Supervised: Regression

- Regression → förutsäg ett tal
- Exempel:
 - Huspris
 - Efterfrågan / försäljning
 - Tid till leverans

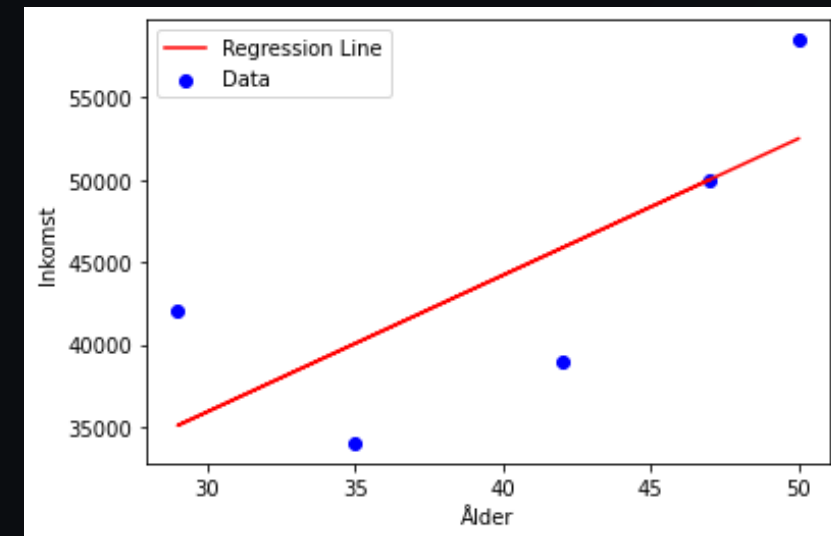
Supervised: Regression

- Den beroende variabeln ($y = \text{inkomst}$) har kontinuerliga värden
- Hur beror inkomst på ålder?

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$$\text{inkomst} = \beta_0 + \beta_1 * \text{ålder} + \varepsilon$$

Inkomst (y)	Ålder (x)
58500	58
42000	29
34000	35
39000	42



Supervised: Klassificering

- Klassificering → förutsäg en klass/etikett
- Binär: Ja/Nej
- Multiklass: Flera kategorier
- Exempel:
 - Spam / inte spam
 - Bedrägeri / ej bedrägeri
 - Sjukdom / frisk

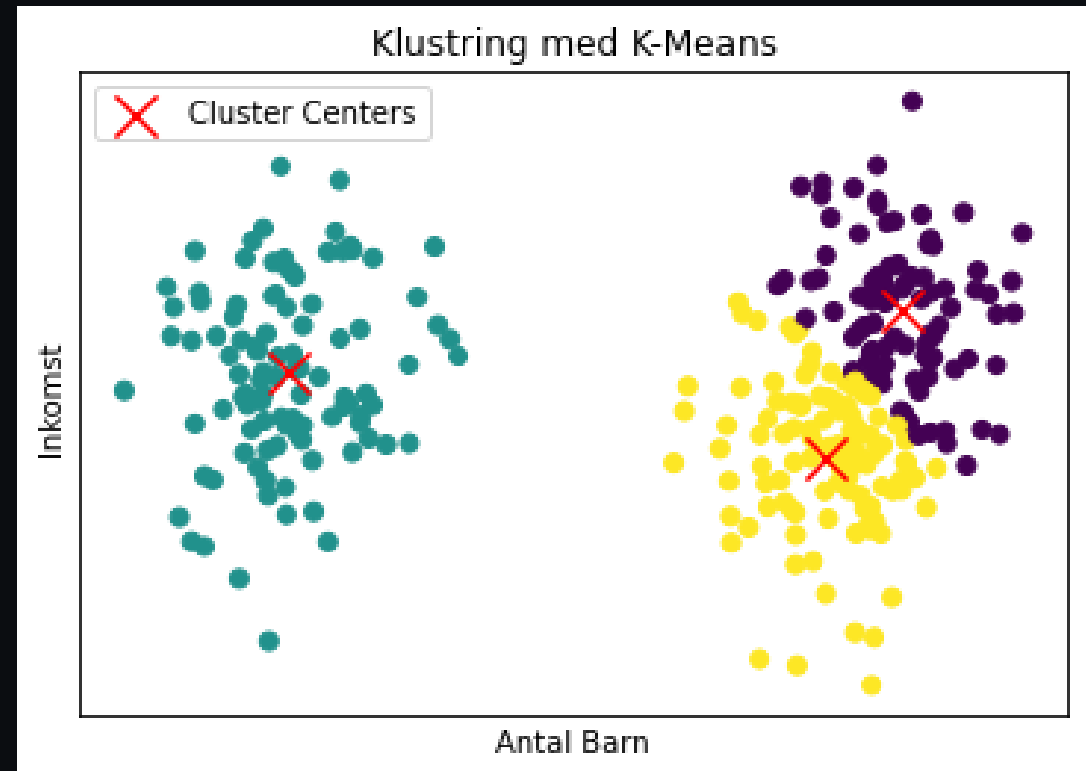
Terminologi

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y = Beroende variabler / Label.
- X = Oberoende variabler / Feature.
- ε = En slumpmässig felterm eftersom sambandet inte är deterministiskt.
- β_0 & β_1 = Parametrar.

Unsupervised: Klustering

- Klustering → hitta grupper i data
- Ingen label (y)
- Exempel:
 - Kundsegment
 - Produktgrupper
 - Avvikande grupper



Unsupervised: Dimensionsreduktion (PCA)

- Dimensionsreduktion → komprimera information
- Syften:
 - Visualisera data
 - Minska brus
 - Snabbare modeller
- Exempel: PCA

Snabbt beslutsträd

Vilken typ är mitt problem?

1. Har jag en beroende variabel (label)?

- Ja → supervised
- Nej → unsupervised

2. Vad är y?

- Tal → regression
- Klass → klassificering

Konst-delen

Modellering handlar om val

- Det finns sällan ett “rätt” svar
- Du måste motivera:
 - Metric
 - Modell
 - Validering
 - Preprocessing

Nästa steg

ML-workflow

Baselines

Train/test