

Programmeerimine vs masinõpe



Masinõpe

- me ei tea (?), kuidas ülesannet lahendada
- meil on näited selle ülesande lahendustest
 - ehk andmed
- me võtame appi üldisema programmi / funktsiooni, mis sobib paljude ülesannete lahendamiseks, kuid vajab mõnede parameetrite häälestamist
- seda saab teha automaatselt!, kasutades andmeid

Masinõpe: mudelid



Mudel: $y = f(x; \theta)$

Masinõpe: mudelid



Mudel: $y = f(x; \theta)$

- x : sisend, y : väljund
 - x : pilt / tekst / isik / idee / ...,
 y : ennustus / hinnang / isik / idee / ...

Masinõpe: mudelid



Mudel: $y = f(x; \theta)$

- x : sisend, y : väljund
 - x : pilt / tekst / isik / idee / ...,
 y : ennustus / hinnang / isik / idee / ...
- f : mudel (nt. programm)
 - algoritm / funktsioon / Turingu masin / reeglistik / ...

Masinõpe: mudelid

Mudel: $y = f(x; \theta)$

- x : sisend, y : väljund
 - x : pilt / tekst / isik / idee / ...,
 y : ennustus / hinnang / isik / idee / ...
- f : mudel (nt. programm)
 - algoritm / funktsioon / Turingu masin / reeglistik / ...
- θ : mudeli parameetrid
 - vektor / maatriks / reeglite nimekiri / ...

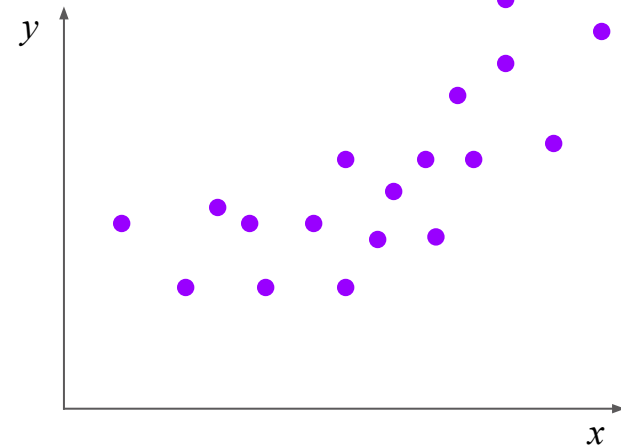
Näide



Korteri suurus x

Kodukindlustuse tasu y

$$y = f(x; \theta) = k \cdot x + b$$

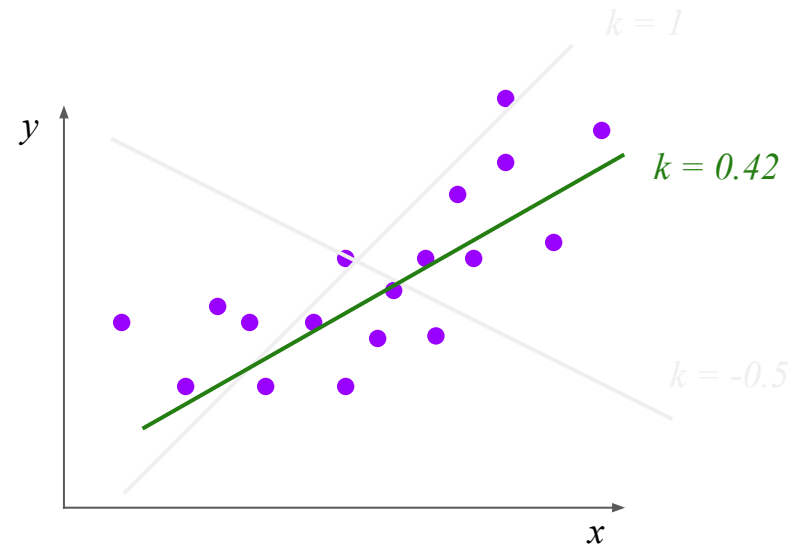


Näide



Korteri suurus x
Kodukindlustuse tasu y

$$y = f(x; \theta) = k \cdot x + b$$



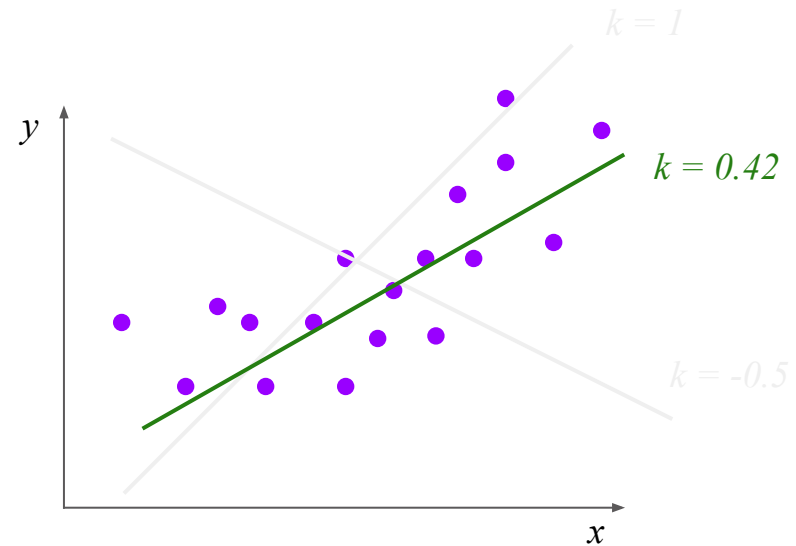
Näide



Korteri suurus x
Kodukindlustuse tasu y

$$y = f(x; \theta) = k \cdot x + b$$

Millised k ja b on head?



Näide

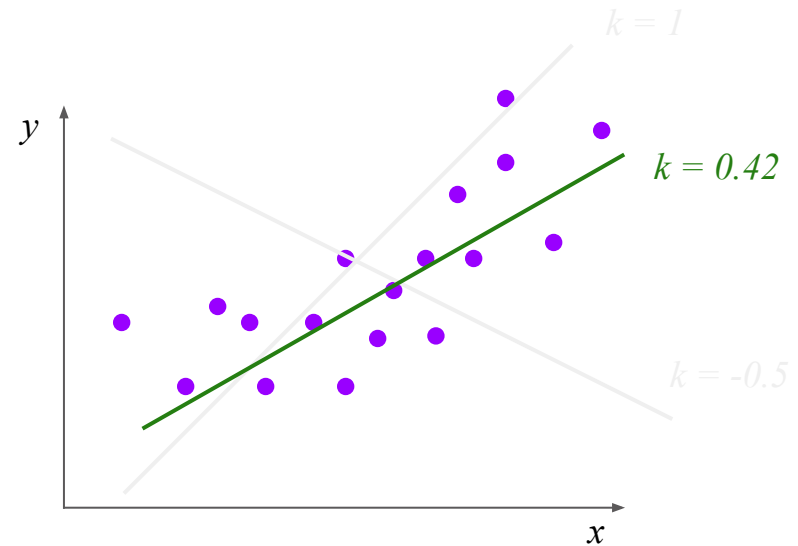


Korteri suurus x
Kodukindlustuse tasu y

$$y = f(x; \theta) = k \cdot x + b$$

Millised k ja b on head?

Vastus: vahe päris y ja $f(x; \theta)$ vahel näitab selle
nt MSE: $\sum_{i=1..n} (y^{(\text{õige})} - y^{(\text{ennustus})})^2$



Tunnused



- **numbriline**
 - diskreetne
 - pidev
 - **väärtused on võrreldavad**

Tunnused



- **numbriline**
 - diskreetne
 - pidev
 - **väärtused on võrreldavad**
- **nominaalne**
 - kahendtunnus (2 väärtust)
 - N väärtust
 - **väärtused pole võrreldavad**

Tunnused



- Tunnuste valik ja sisu olenevad sisendobjektidest ning lahendatavast ülesandest (või kättesaadavatest andmetest)
- NB! Kasulike tunnuste valimine on kõige olulisem osa masinõppe kasutamisest, kuna see mõjutab tulemusi kõige rohkem