→ Gjøres sammen med PowerPoint "Forelesning 1.1"

Forelesning 1 - Informasjon.

▸ Velkommen til SOK-3023!
"Maskinlæring for økonomer."

↳ Introduksjon av meg selv. → Markus J. Aase
→ Bakgrunn i matematikk
→ Master om ML / sneskred. Og bruk nye modeller i økonomiske anvendelser.

↳ Vi vil dekke "bittelitt" av et stort fagfelt, kalt maskinlæring / kunstig intelligens. I et fagfelt som utvikler seg hurtig, dag for dag.

↳ Dette kurset vil ha en matematisk / statistisk grunnlag, mer enn mange av deres andre kurs - MEN, vi vil ha hovedfokus på ~~andre~~ anvendelser, prosjektarbeid, forståelse.

↳ Kurset er lagt opp følgende

- Prosjektarbeid - 50%
  - ↳ Presentasjon av prosjektbeskrivelse
  - ↳ Gr. str. 1-2
- Muntlig eksamen - 50%
  - ↳ Individuell

"Noe av det som er spennende med dette fagfeltet er at det går så hurtig framover."
"Jeg vet ikke om dette kurset vil se likt ut,

Hva er intelligens?

- Jeg liker å tenke på intelligens som:

"Er evnen til å prosessere informasjon, som vil gi kunnskap til ~~[illegible]~~ å ta beslutninger i fremtiden."

↳ Dette gjør vi hver eneste dag. Fra å knyte sko, sjekke trafikken før vi krysser gata osv.

↳ AI, er at vi "gir" datamaskiner muligheten til å gjøre det samme. Og dette gjøres ved hjelp av <u>DATA</u>

Kurset i én setning:

Lære datamaskiner hvordan lære en oppgave fra innsamlet data.

Introduksjon til maskinlæring - del 1

- Statistikk
- Kalkulus
- Lineær algebra.

Nødvendig lineær algebra

"Lineær algebra er den delen av matematikken som omhandler vektorer og vektorrom."

Brukes i computer vision, maskinlæring, signalprosessering og alt rundt oss.

Vi vil fokusere på noen få konsepter.

Vektor:

En vektor er en ordnet liste av tall som kan representere datapunkter i et flerdimensjonalt rom.

En vektor $\mathbf{x}$ i n-dimensjoner:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

x_i er komponentene i vektoren.

Matrise

Rektangulær array/skjema av tall (symboler), og brukes ofte til å representere data eller transformasjoner.

$$A = \begin{pmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & \ddots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

dimensjon $m \times n$

horisontalt = rader

vertikale = kolonner

"Både når det gjelder matriser og vektor er flere operasjoner viktig i maskinlæring."

" - Transponering
- Skalar multiplikasjon
- Matrisemultiplikasjon
- ... "

Eksempel: Forutsi lønn basert på utdanning erfaring.

$$\mathbf{X} = \begin{bmatrix} 12 & 3 \\ 16 & 5 \\ 14 & 2 \\ 10 & 0 \\ \vdots & \vdots \end{bmatrix} \qquad Y = \begin{bmatrix} 400 \\ 600 \\ 500 \\ 350 \\ \vdots \end{bmatrix}$$

utdanning erfaring — $m \times 2$

lønn — m

$$Y = X \cdot \beta + \varepsilon, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$\hat{Y} = X \cdot \hat{\beta}$$

Hvorfor?

I kurset skal vi bruke Tensorflow.

Tensorflow

- optimalisert for beregning av matriser og tensorer.
- Når data lagres i vektor/matriseformat, muliggjør det parallellbehandling.
  - "Innebærer at flere operasjoner kan utføres samtidig, noe som utnytter kraften til datamaskiner som GPU og fler-kjernet CPU."

Tensorer

"En "vektor" (array) av tall som er plassert i en grid med variabelt antall av akser!"

1. **Skalar** (rank-0 tensor) $\mathbb{R}$
   - Tallet 3, 7.5, osv.
   - Matematikken: kun har størrelse, ikke retning.

2. **Vektor** (rank-1 tensor) $\mathbb{R}^n$
   - Har størrelse og en retning

$$\begin{bmatrix} 1 \\ 3 \\ 5 \\ 7.5 \end{bmatrix}$$

3. **Matrise** (rank-2 tensor) — viktig å kunne regne med! $\mathbb{R}^{n \times m}$

$$\begin{bmatrix} 3 & 4 & 8 \\ 2 & 6 & 9 \\ 1 & -2 & 15 \end{bmatrix}$$

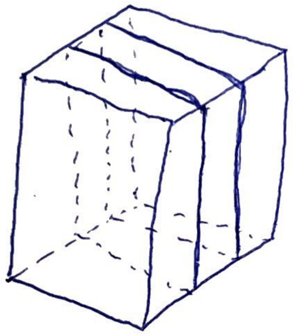
4. Rank-3 tensor

↳ "Som en kube"
↳ Tenk på Rubiks kube

$$\begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 1 \\ 8 & 4 & 9 \end{bmatrix} \begin{bmatrix} 4 & 2 & 6 \\ 8 & 5 & 9 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 2 \\ 8 & 9 & 6 \end{bmatrix}$$

$$\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \quad (\mathbb{R}^{n \times m \times p})$$ I ML/AI

RGB = fargebilde

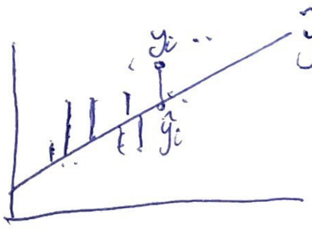


"representerer"
Rank 3 tensor = fargebilde
Rank 2 tensor = Sort/hvitt

Normer

"Viktig konsept i lin.alg."

Normen til en vektor **x** er en måling av dens lengde.



Generelle normer er definert

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Eks: L1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

L2-norm $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

Dette relaterer til MSE

$r = y_i - \hat{y}_i$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MSE = \frac{\|r\|_2^2}{n}$$

"Normer brukes i optimalisering for å evaluere og minimere tap."
Mer om dette senere.

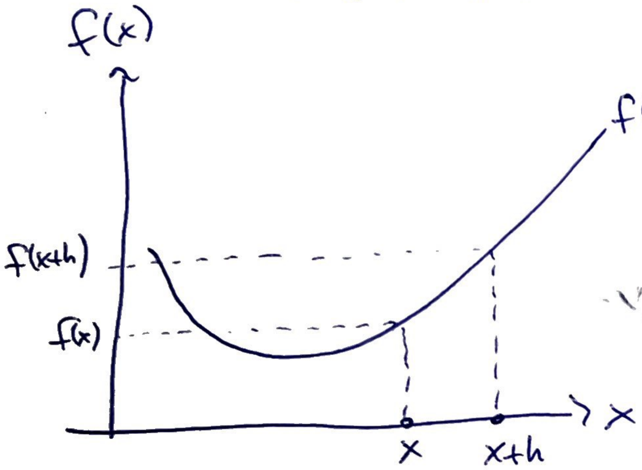
Kalkulus

"Litt kalkulus og derivasjon er avgjørende for å forstå hvordan maskinlæringsmodeller blir trent og optimalisert.

Derivasjon:

"betyr latinske Limes = grense"

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



"Brukes til å finne stigningstallet til tangenten til funksjonen f(x) i punktet x."

Da kan vi finne hvor en funksjon minker, og vi kaller det en tapsfunksjon eller kostfunksjon.

Irreducible versus Reducible error

Statistikk

Forventningsverdi:

$$E[X] = \sum_x x \cdot f(x) \text{, hvis } X \text{ er diskret}$$

~~Varians:~~ $$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx \text{, hvis } X \text{ er kont.}$$

Varians:

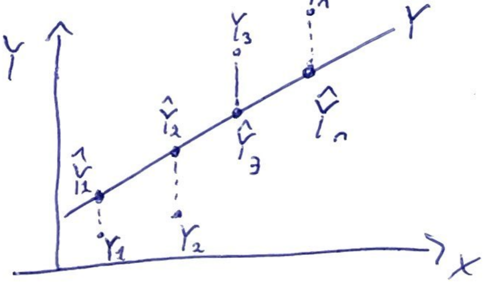
X er en tilfeldig variabel med ss. ford. $f(x)$

(Diskret) $$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 \cdot f(x)$$

(Kont.) $$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$



Vi har observert respons Y og predikert respons $\hat{Y}$

- Vi ønsker å se på forskjellen her.

$$Y = f(X) + \varepsilon \quad , \quad \hat{Y} = \hat{f}(X)$$

$$E[(Y-\hat{Y})^2] = E\left[\left(f(X)+\varepsilon-\hat{f}(X)\right)^2\right]$$

$$= E\left[\left(f(X)+\varepsilon-\hat{f}(X)\right)\left(f(X)+\varepsilon-\hat{f}(X)\right)\right]$$

$$= E\left[\left(f(X)-\hat{f}(X)\right)\left(f(X)+\varepsilon-\hat{f}(X)\right)+\varepsilon\left(f(X)+\varepsilon-\hat{f}(X)\right)\right]$$

$$= E\left[\left(f(X)-\hat{f}(X)\right)^2+\varepsilon\left(f(X)-\hat{f}(X)\right)+\varepsilon\left(f(X)-\hat{f}(X)\right)+\varepsilon^2\right]$$

Forventningsverdi er lineær! $$\left(E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]\right)$$

$$E[(f(x)-\hat{f}(x))^2] + E[\varepsilon^2] + 2E[\varepsilon(f(x)-\hat{f}(x))]$$

Forventningsverdien til f og $\hat{f}$ er konstant

$$= [f(x)-\hat{f}(x)]^2 + E[\varepsilon^2] + 2E[\varepsilon(f(x)-\hat{f}(x))]$$

Husk at $\varepsilon \sim N(0, \sigma^2) \Rightarrow E[\varepsilon] = 0$

$$= [f(x)-\hat{f}(x)]^2 + E[\varepsilon^2]$$

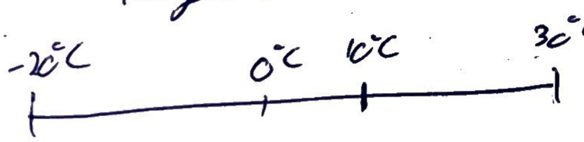
Når $E[\varepsilon] = 0 \Rightarrow E[\varepsilon^2] = Var(\varepsilon)$

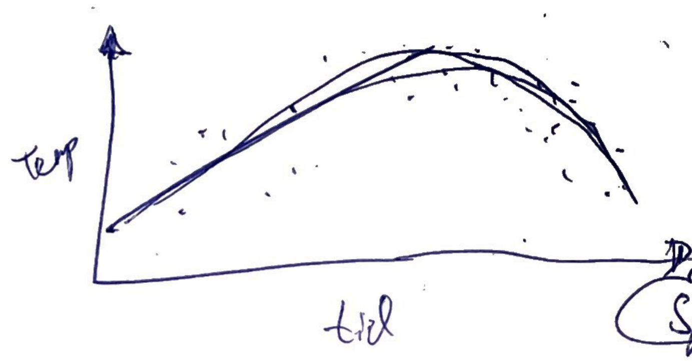
$$= \underbrace{[f(x)-\hat{f}(x)]^2}_{\text{reducible error}} + \underbrace{Var(\varepsilon)}_{\text{irreducible}}$$

Regresjon versus **klassifikasjon**

- Kontinuerlig variabel (Output)

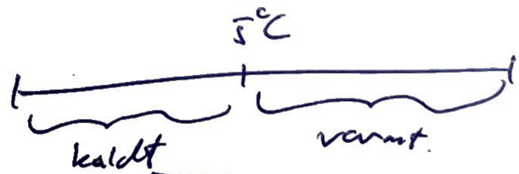
"Hva er temperaturen imorgen?"

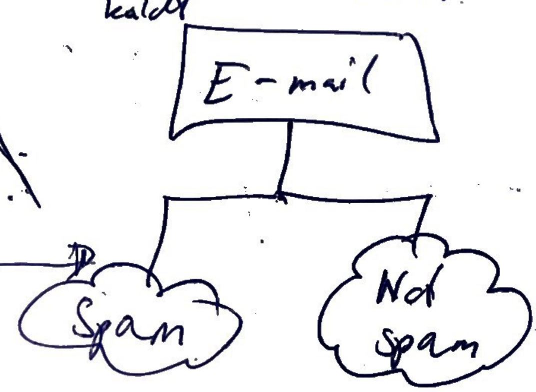




- Kategorisk variabel (Output)

"Er det kaldt eller varmt imorgen?"





... kort om veiledet læring / ikke-veiledet læring

Prediksjon vs Inferens

Veldig viktig å vite forskjellen her ⚠

Vi tar et eksempel med lineær regresjon

- Kvantitativ respons Y (kalles og output, målvariabel / targetvariabel, avhengig variabel.)
- p ulike prediktorer

$$X = (X_1, X_2, \dots, X_p)$$

(kalles og uavh. variabel, forklarende -"-, input, kovariater, features)

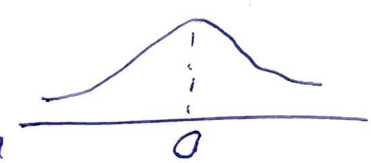
(Her kan X være matrise og Y responsene)

Hvis vi har multippel lineær regresjon, er sammenhengen (tenkt) lineær mellom Y og X.

$$Y = f(X) + \varepsilon$$

ukjent funksjon

$$\varepsilon \sim N(0, \sigma^2)$$



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

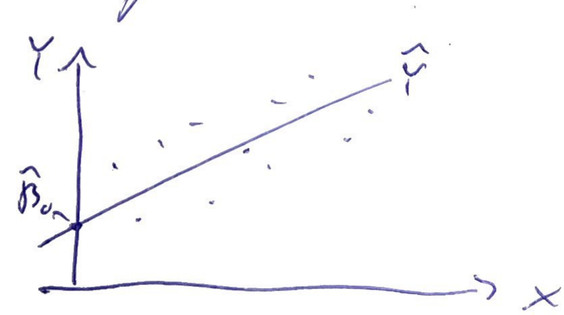
Prediksjon:

$$\hat{Y} = \hat{f}(X)$$
$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p$$

$$\boxed{E[\varepsilon] = 0 \ ⚠}$$

"Multippel lin. reg. er vanskelig å visualisere, men hvis man har enkel lin. reg"

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$$



Prediksjon

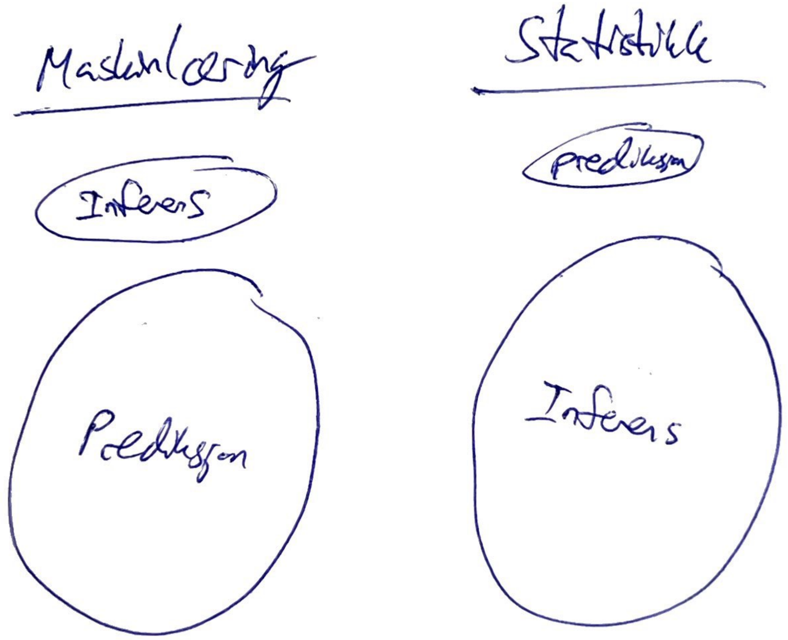
- Prosessen å forutsi fremtidige utfall basert på tilgjengelig data.
- Black box

$$Y = \hat{f}(X)$$

black box

Inferens

- Fokuserer på å trekke/prøver å konklusjoner om ~~de underliggende prosessene~~ egenskapene til en underliggende sannsynlighets-fordeling. Finne relasjon mellom output av koeffisienter.
- Konfidens intervall, hypotesetesting, t-tester osv.



Kodelab #1

- Start med kort om veiledet læring / ikke-veiledet læring
- Ila dato: evaluering av klassifikasjonsmodeller versus regresjonsmodeller

- Git og henting av notebooks — Vis Google Colab interface
- Scikit learn
- data = fetch_california_housing
  data['DESCR'] ← Se på denne !!

pd.DataFrame(data.data , columns=data.feature_names) ↙ Matrise

~~pd.DataFrame(~~
pd.Series(data.target, name='MedianHouseValue') ↙ Vektor

- Evaluering av modellen. Ta med

  MAE = mean_absolute_error

Enkel versus Mult. linear regresjon < MSE minker / R^2 øker

Iris

- Snakk om label_encoder av species
- Accuracy, confusion matrix
-