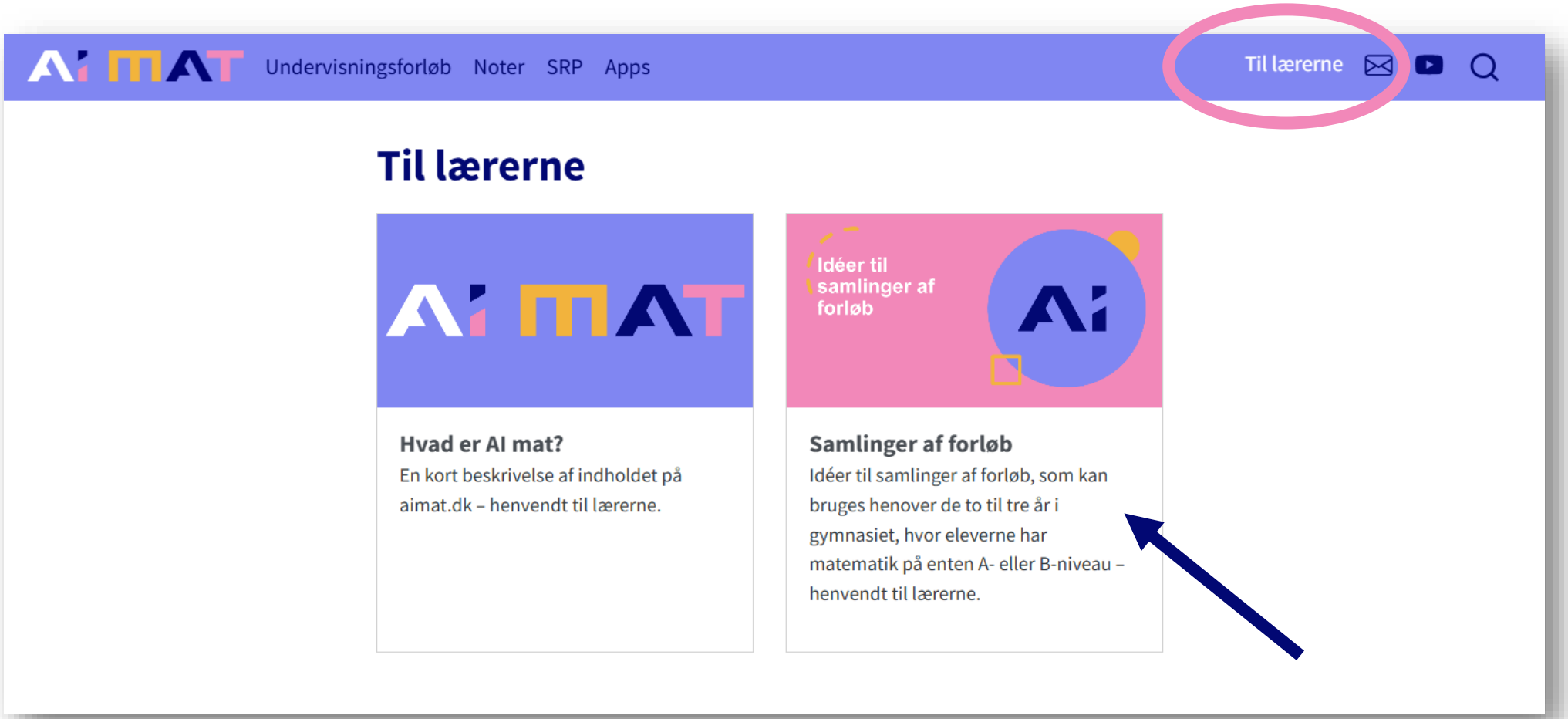
AI MAT

MATEMATIKKEN BAG MAGIEN

Ege Rubak / Institut for matematiske fag, AAU
Malene Cramer Engebjerg / Aalborg Katedralskole

Idéer til forløb



Forslag 5

AI MAT
MATEMATIKKEN BAG MAGIEN

Hvornår?	Omfang	Forudsætninger	Forløb
1g	1,5-2 timer	Koordinatsystemer, punkter og afstand mellem punkter, procentregning	Overvågning i Monitorbian
1g	2 timer	Deskriptiv statistik, koordinatsystemer, rette linjer	Kan vi genkende håndskrevne tal?
2g	1-2 timer	Polynomiel regression	Overfitting og krydsvalidering med polynomiel regression
2g	1,5-2 timer	Vektorregning, differentialregning	Word2vec
3g	6-9 timer	Sammensatte funktioner, differentialregning	Kunstige neuroner (inkl. forslag til mundtlig eksamensspørgsmål)

Overvågning i Monitorbian

Landet *Monitorbian* ønsker at blive en vaskeægte overvågningsstat! Men efterretningstjenesten i Monitorbian ved meget lidt om overvågning. Derfor har de ansat jer som intelligence officerer til at løse denne opgave. Tillykke med jeres nye job! Lad os smøge ærmerne op og komme i gang! 😀

Forudsætninger og tidsforbrug

Forløbet kræver kendskab til:

- Koordinatsystemer.
- Punkter og afstande mellem punkter.
- Procentregning.

Tidsforbrug: Ca. 90 minutter (uden brug af Orange).

C-niveau

Overvågning i Monitorbian

Eleverne får bekendtskab til

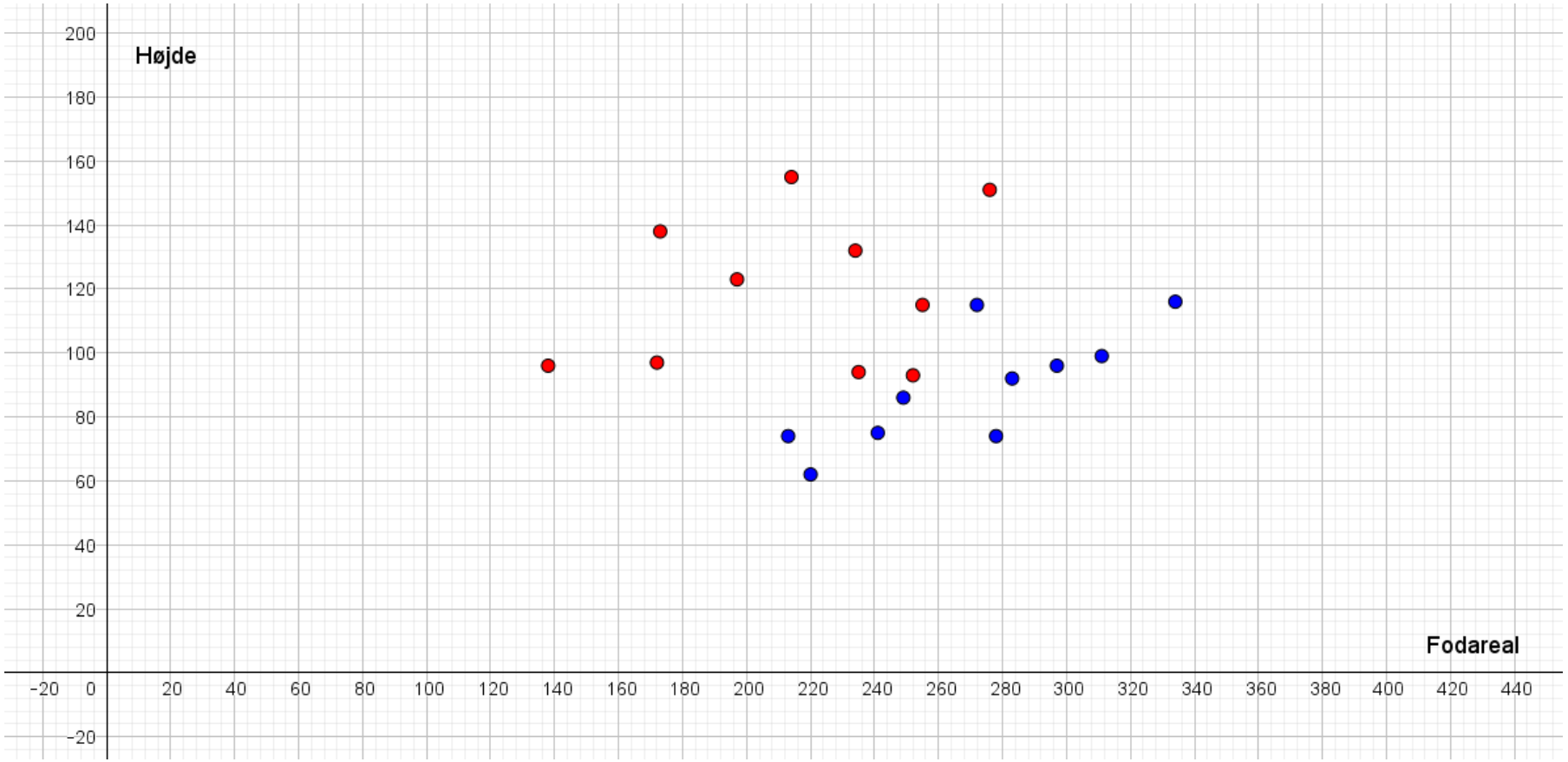
- Features
- Targetværdier
- Træningsdata
- Testdata
- Confusion matrix
- Klassifikationsnøjagtighed



Overvågning i Monitorbian







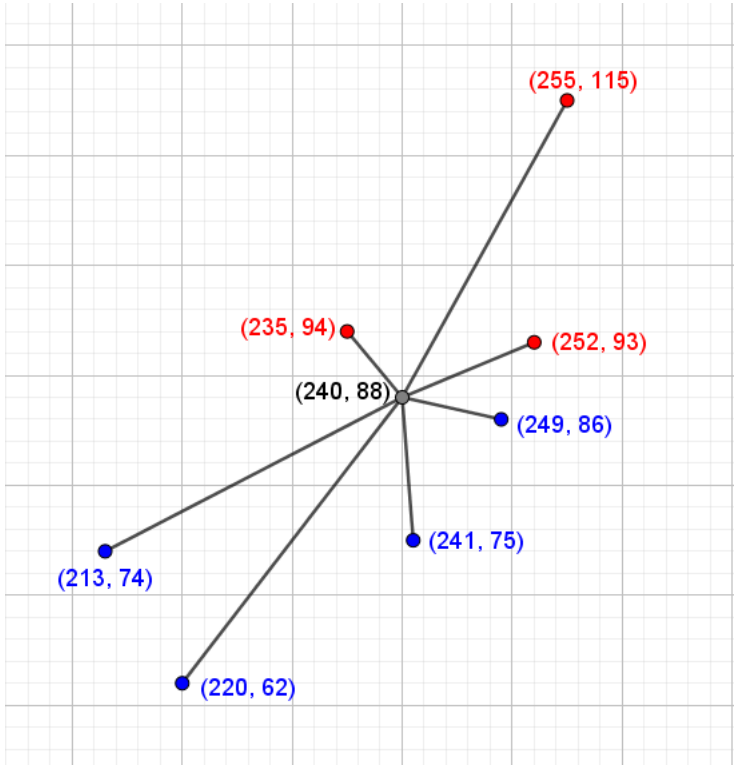
Overvågning i Monitorbian

AI-metode

- Nærmeste naboer (kNN)

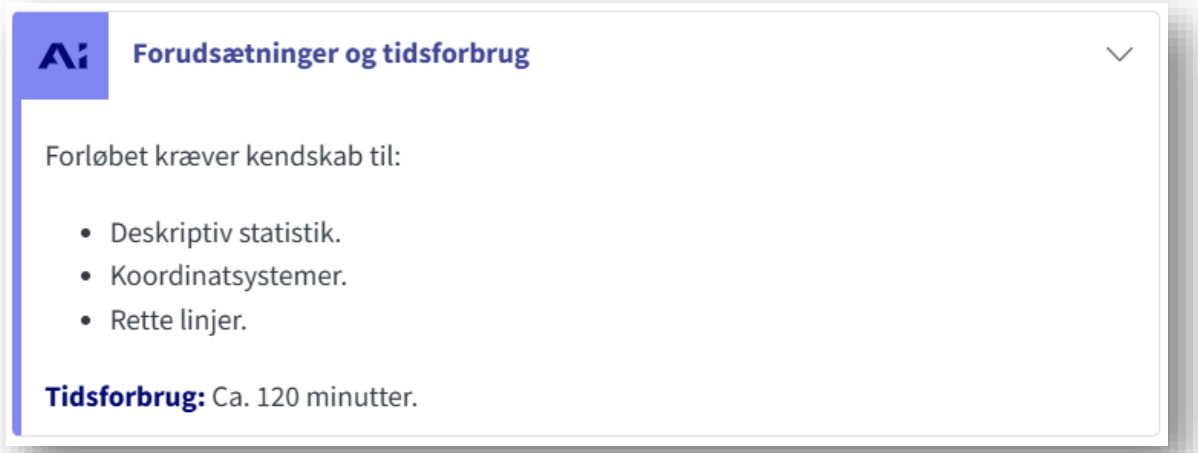
- Hvordan vælges k vha. testdata?

Til sidst: kNN direkte på billederne i Orange



Kan vi genkende håndskrevne tal?

Nogle gange har man brug for, at håndskrevet tekst kan digitaliseres. Det kan for eksempel være i en eksamenssituation, hvor din lærer og censor skriver din karakter på en liste, som efterfølgende skal tastes ind på en computer. Det ville være smart, hvis man kunne scanne listen eller tage et billede af den, hvorefter alle tal korrekt vil blive genkendt (så du ikke får en forkert karakter!).

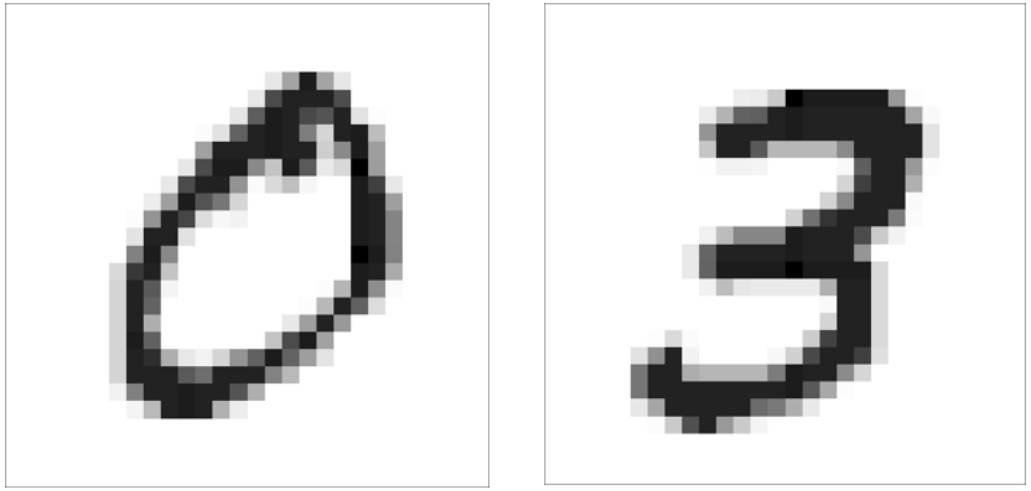


C-niveau

Kan vi genkende håndskrevne tal?

Eleverne får kendskab til

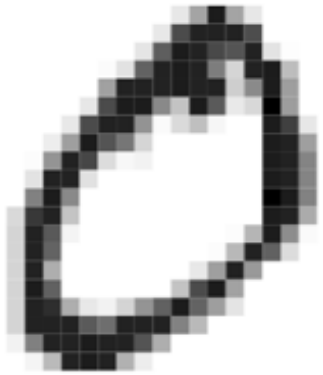
- Hvordan billeder repræsenteres på en computer
- Features
- Targetværdier
- Træningsdata
- Testdata
- Confusion matrix
- Klassifikationsnøjagtighed
- Feature engineering



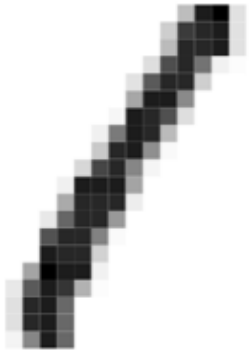
MNIST datasættet

Kan vi genkende håndskrevne tal?

- Kan vi se forskel på 0 og 1?



0	0	0	0	0	0	0
0	0	1	104	211	27	0
0	0	105	164	96	129	0
0	36	132	3	0	172	0
0	84	53	7	114	47	0
0	58	207	156	36	0	0
0	0	0	0	0	0	0



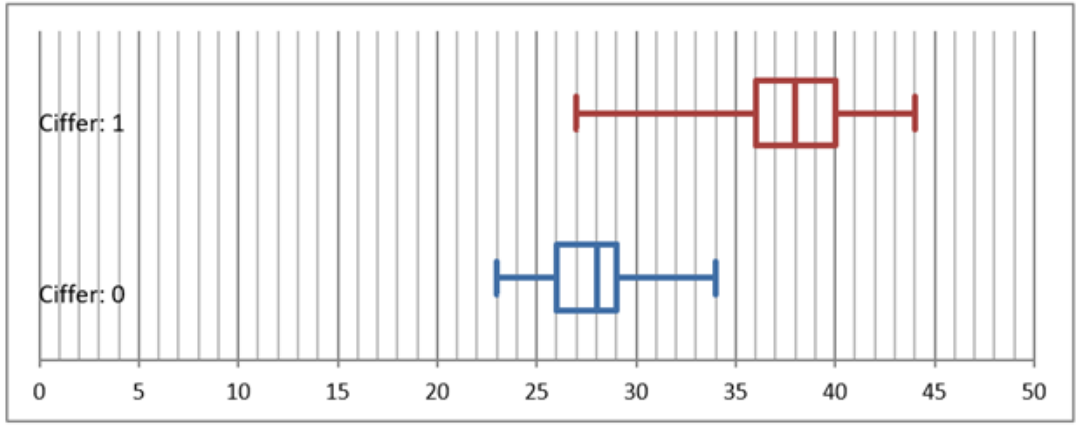
0	0	0	0	0	0	0
0	0	0	0	100	59	0
0	0	0	30	177	2	0
0	0	2	174	37	0	0
0	0	88	142	0	0	0
0	0	189	27	0	0	0
0	0	43	0	0	0	0

- Beregning af features: antallet af hvide pixels, den gennemsnitlige pixelværdi

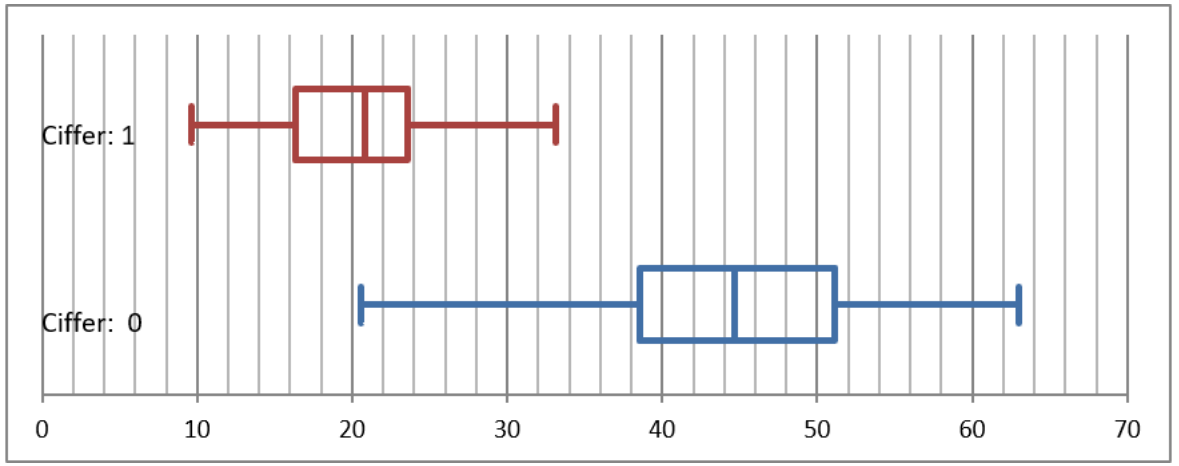
Kan vi genkende håndskrevne tal?

AI-"metode" (det behøver ikke være langhåret)

- Vurdering vha. boksplots



Antal tomme pixels

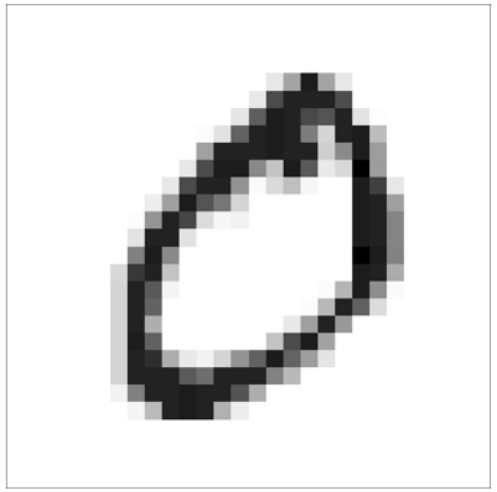


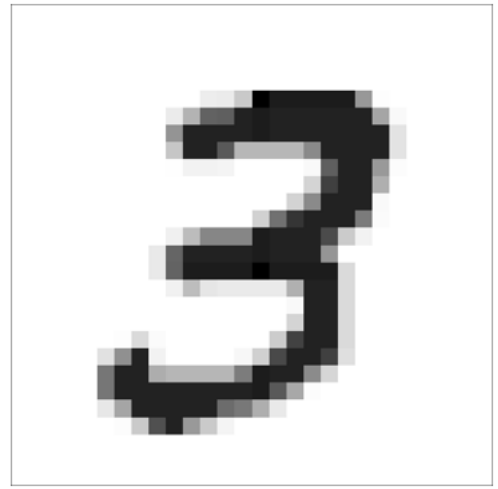
Gennemsnitlig pixelværdi

Kan vi genkende håndskrevne tal?

- Ved alene at tælle antal hvide pixels kan vi kende forskel på 0- og 1-taller med en klassifikationsnøjagtighed på 98.5%

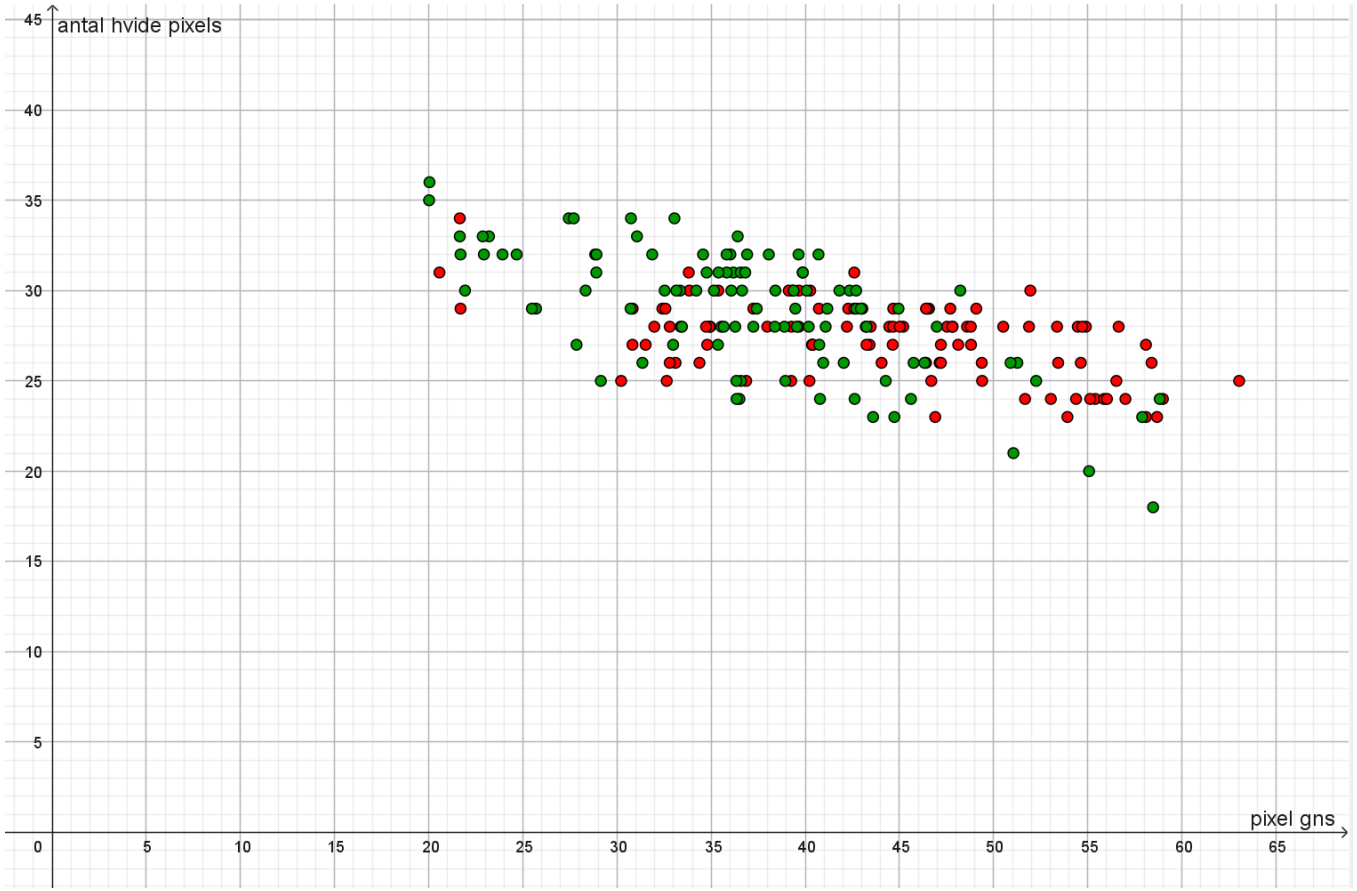
- Hvad med 0 og 3?

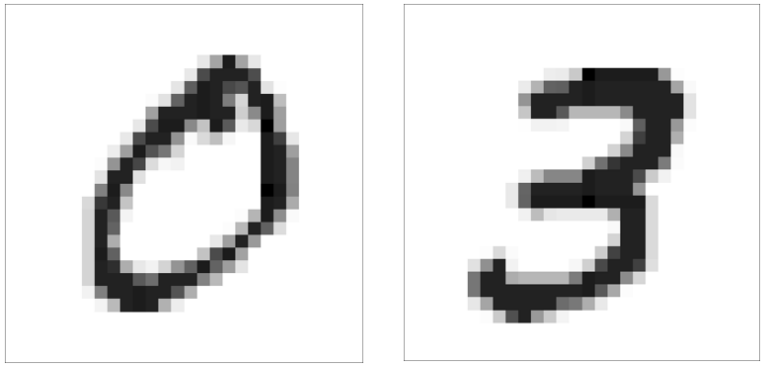




Kan vi genkende håndskrevne tal?

- De samme features dur ikke til 0 og 3...





Kan vi genkende håndskrevne tal?

- Nye features: Gennemsnitlige vandrette og lodrette lokale maxima
  (vi ser bort fra de rækker, hvor der ikke er nogle maxima)

0	0	0	0	0	0	0
0	0	1	104	211	27	0
0	0	105	164	96	129	0
0	36	132	3	0	172	0
0	84	53	7	114	47	0
0	58	207	156	36	0	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	0	70	166	189	73	0
0	0	42	41	167	104	0
0	0	118	201	210	15	0
0	5	13	17	175	0	0
0	149	162	175	94	0	0
0	7	48	2	0	0	0

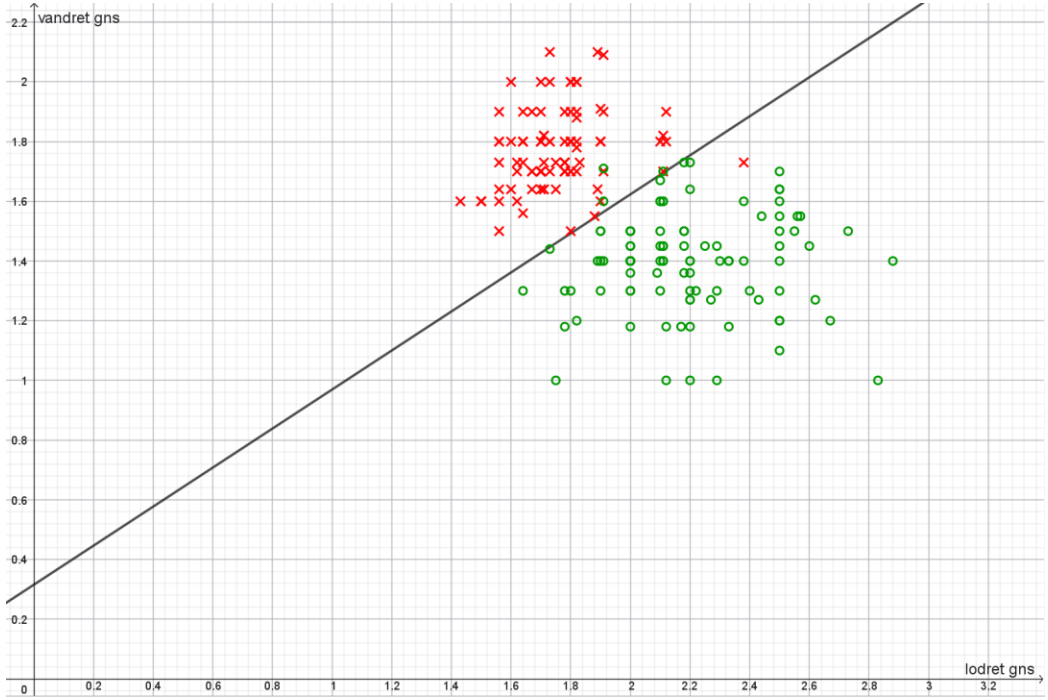
0	0	0	0	0	0	0
0	0	1	104	211	27	0
0	0	105	164	96	129	0
0	36	132	3	0	172	0
0	84	53	7	114	47	0
0	58	207	156	36	0	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	0	70	166	189	73	0
0	0	42	41	167	104	0
0	0	118	201	210	15	0
0	5	13	17	175	0	0
0	149	162	175	94	0	0
0	7	48	2	0	0	0

Kan vi genkende håndskrevne tal?

AI-"metode"

- Tegn selv en ret linje, som kan adskille to grupper af punkter



Pixel-opløsning: 14

Klassifikationsnøjagtighed: 98%

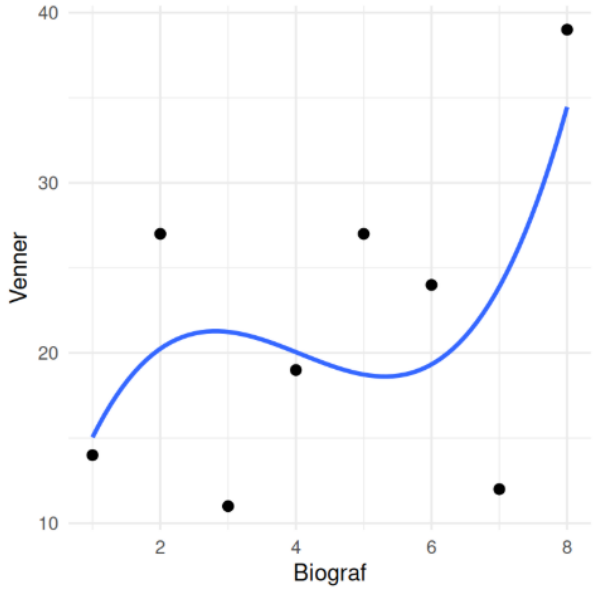
Kan vi genkende håndskrevne tal?

App



Overfitting og krydsvalidering med polynomiel regression

- Forudsætning: Polynomiel regression
- Datasæt



3. gradsregression

Antal biografbesøg det seneste år	Antal venner på de sociale medier
1	14
2	27
3	11
4	19
5	27
6	24
7	12
8	39

B-niveau

Overfitting og krydsvalidering med polynomiel regression

- Eleverne laver 2. – 7. grads regression på data

- Men er 7. grads polynomiet godt til at prædiktere fremtidige observationer?

Overfitting og krydsvalidering med polynomiel regression

- Hands-on 4-folds krydsvalidering

Antal biografbesøg det seneste år	Antal venner på de sociale medier
1	14
2	27
3	11
4	19
5	27
6	24
7	12
8	39

Testdata

Træningsdata

Antal biografbesøg det seneste år	Antal venner på de sociale medier
1	14
2	27
3	11
4	19
5	27
6	24
7	12
8	39

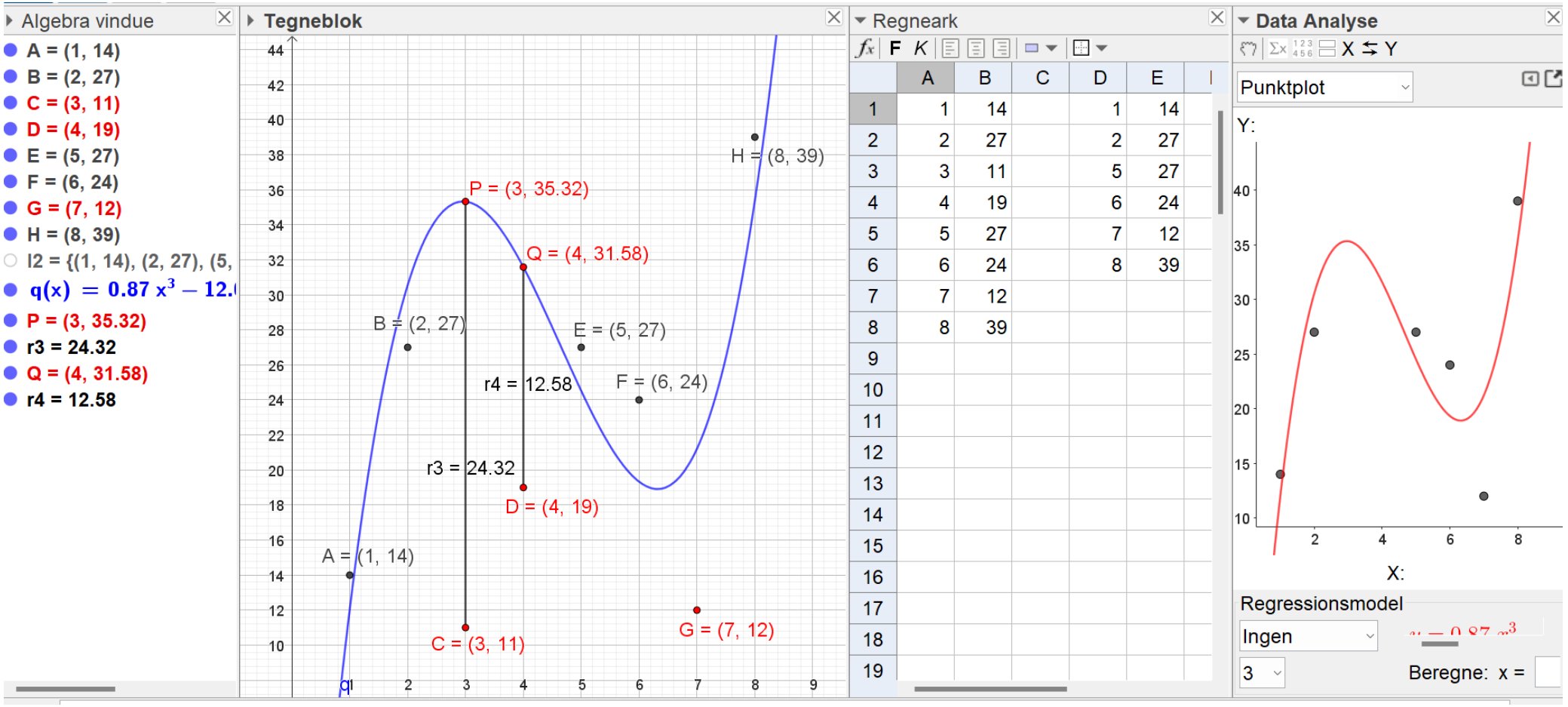
Træningsdata

Testdata

Træningsdata

osv.

Overfitting og krydsvalidering med polynomiel regression – i GeoGebra



Overfitting og krydsvalidering med polynomiel regression

- På baggrund af 3. grads regression finder eleverne residualer:

Residual	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
Værdi								

- Herefter udregner de summen af de kvadrerede residualer:

$$r_1^2 + r_2^2 + \cdots + r_8^2$$

Overfitting og krydsvalidering med polynomiel regression

- Eleverne sammenligner summen af de kvadrerede residualer for 3. grads polynomiet med:

Grad	Summen af de kvadrerede residualer
1	146513
2	142952
4	19821534237
5	66277073433

Det lyder som dig!

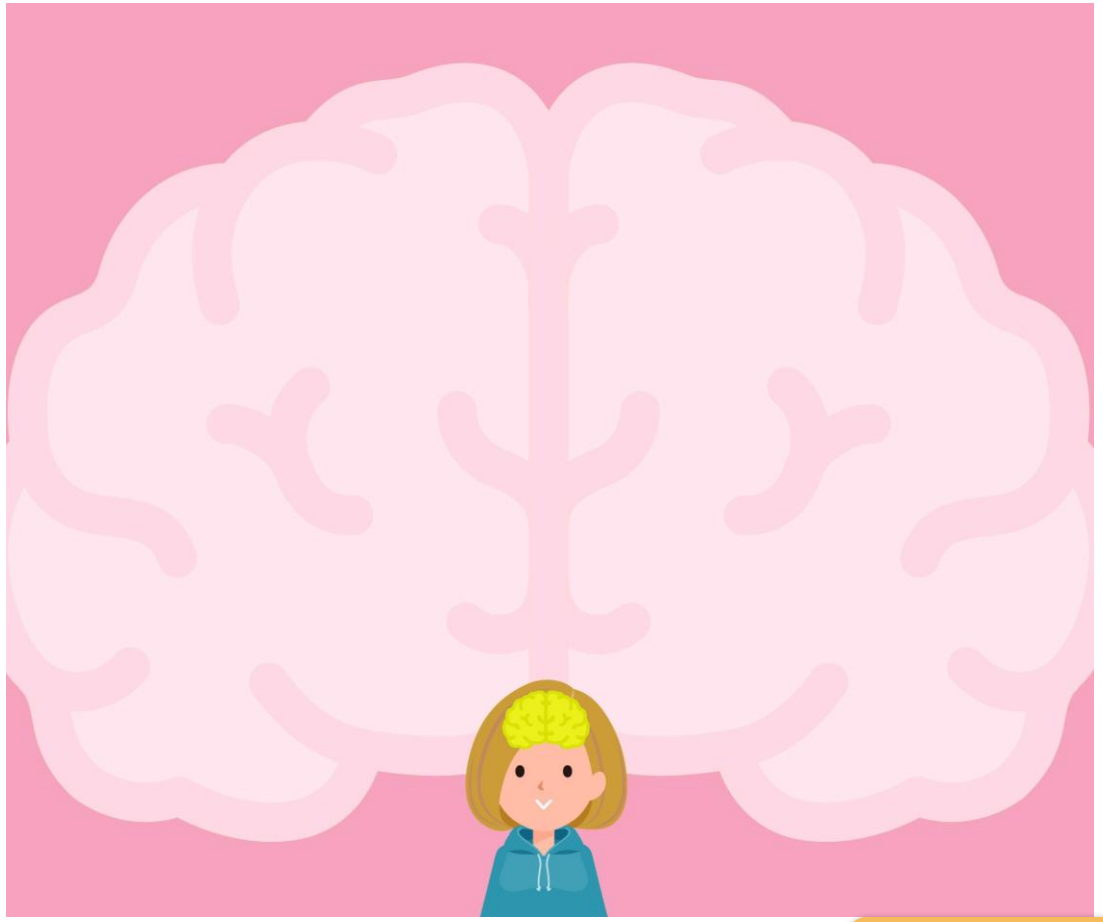
- Ikke så meget matematik, men eleverne stifter bekendtskab med begreber som:
  - Træningsdata
  - Testdata
  - Klassifikationsnøjagtighed
  - Confusion matrix
- Databehandling i Orange



C-niveau

Cambridge Analytica skandalen

- Samarbejde med informatik
- Kan vi prædiktere folks personlighed på baggrund af Facebook-data? (her graden af "åbenhed" (Openness))
- AI-metode: Kunstig neuron



A-niveau

Cambridge Analytica skandalen

- Informatik: Implementation af en app, hvor man kan tage en OCEAN-personlighedstest

- Ekstra udfordring med brug af matematik:
  - Standardisering af OCEAN-scoren vha. normalfordelingen

SRP?



Inddeling af musik i grupper

- AI-metode: K-means (**unsupervised learning**)



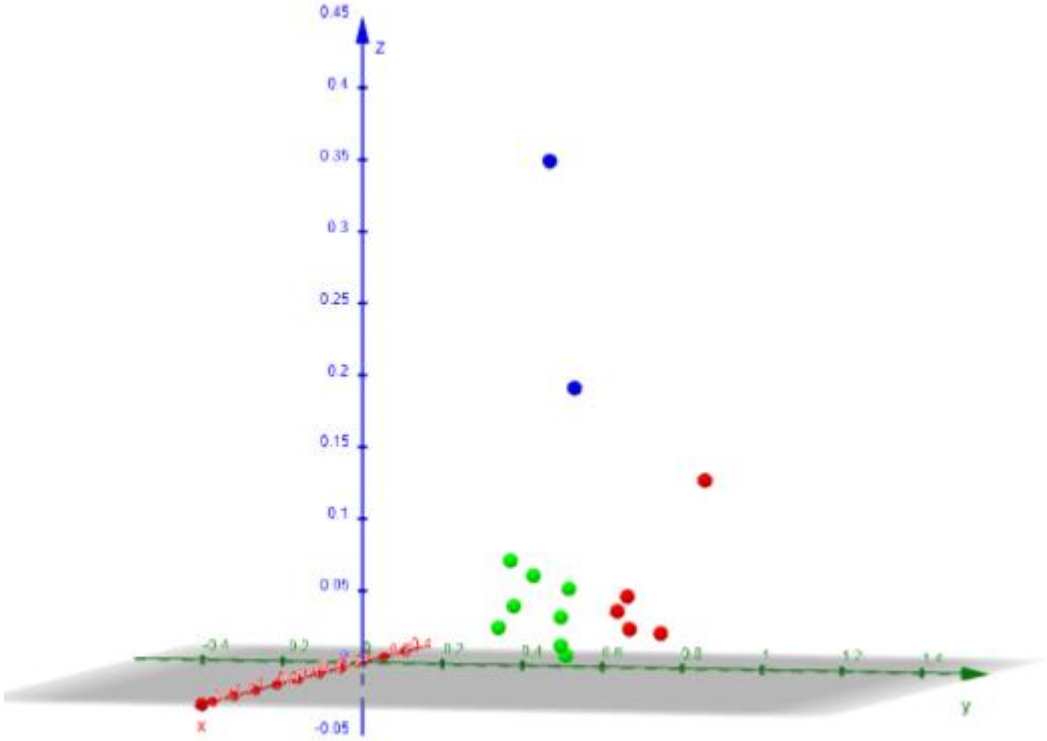
Nummer	Danceability	Energy	Speechiness
Barbie Girl	0.76	0.95	0.04
Shape of You	0.83	0.65	0.08
Set Fire to the Rain	0.60	0.67	0.02
Wake Me Up	0.53	0.78	0.05
Baby One More Time	0.76	0.70	0.03
Just The Way You Are	0.64	0.84	0.04
New Rules	0.76	0.72	0.07
Stan	0.78	0.74	

B-niveau

Inddeling af musik i grupper

- Eleverne regner selv – herefter brug af Orange eller Excel





Hvilken politiker er du mest enig med?

- Simpel kandidattest
- Forudsætning: afstand mellem to punkter
- Både euklidsk afstand og Manhattan afstand



C-niveau

Virker medicinen?

- AI-metode: Support Vector Machines

- Forudsætninger:
  - Linjens ligning: $ax + by + c = 0$
  - Planens ligning
  - Afstand fra punkt til linje
  - Afstand fra punkt til plan
  - Cirklens ligning



A-niveau

Virker medicinen?



- Bevis for

Sætning 1

Hvis $az_1 + bz_2 + c > 0$, så ligger punktet $P = (z_1, z_2)$ på den side af linjen, som $\vec{n} = \begin{pmatrix} a \\ b \end{pmatrix}$ peger mod, og hvis $az_1 + bz_2 + c < 0$, så ligger punktet på modsat side.

- Support vector machines algoritmen implementeres i Excel (der er videoer, som viser hvordan)
- Support vector machines i 3 dimensioner
- Ikke-lineære support vector machines

SRP?

AI, klimaudfordringer og sammenstyrtede broer

AI MAT
MATEMATIKKEN BAG MAGIEN

- Forudsætninger: Lineær regression
- Modellering af betons trykstyrke
- Eleverne laver:
  - Multipel lineær regression på testdata
  - Modeller vurderes ved at lave (y_{data}, y_{model})-plots
  - App bruges til at lave kunstigt neuralt netværk



A-niveau