

Kunstig intelligens i hjerneforskning

Esten H. Leonardsen

19.10.23

Plan for dagen

1. Historien bak kunstig intelligens (AI)
2. Terminologi og konsepter
3. Teori
 - Hva er en maskinlæringsmodell?
 - Hvordan trenne en maskinlæringsmodell?
 - Intuisjon bak bildegenkjenningssystemer
4. AI i hjernehjerneforskning
 - Segmentering
 - Diagnostisk klassifikasjon
 - Hjernealder
 - Forklarbar kunstig intelligens
 - Genererte hjernebilder
5. Forskningsetiske spørsmål ved bruk av AI
6. Generelle etiske spørsmål ved bruk av AI

Historien bak kunstig intelligens

Historien bak kunstig intelligens

Turing-
testen
(1950)



Alan Turing

Historien bak kunstig intelligens

Turing-
testen
(1950)

M I N D

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

By A. M. TURING

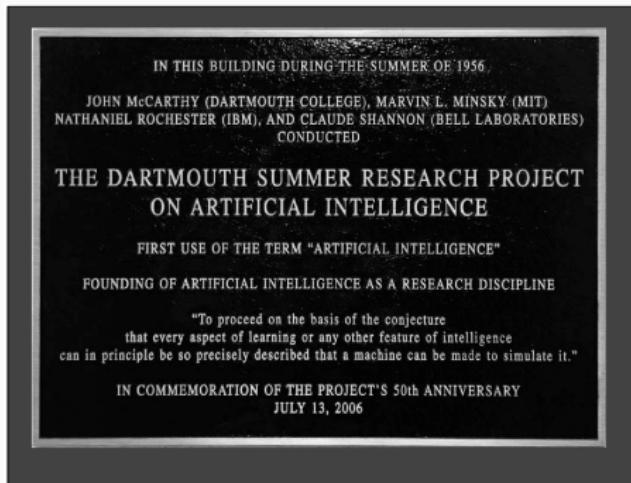
1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think ?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think ?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

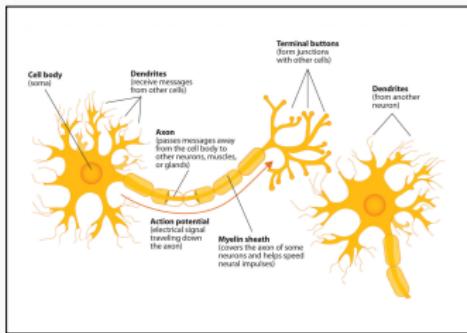
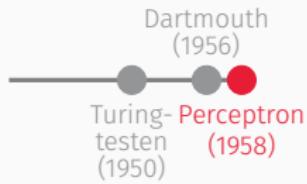
Historien bak kunstig intelligens

Dartmouth
(1956)

Turing-
testen
(1950)

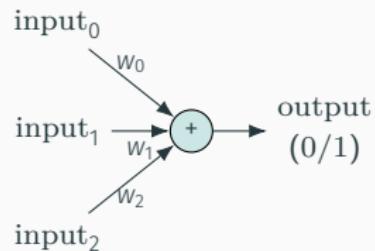
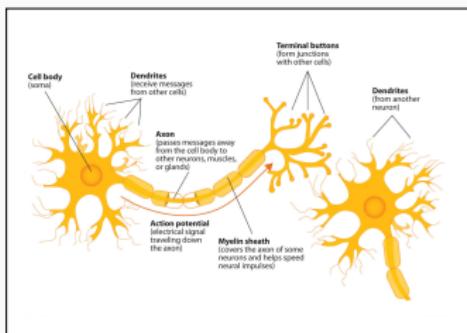
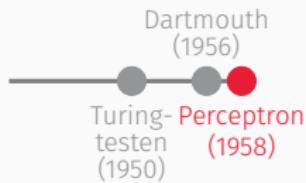


Historien bak kunstig intelligens

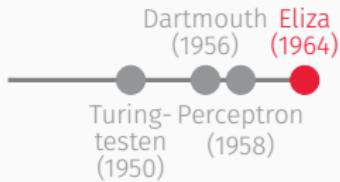


The neuron is the building block of the nervous system

Historien bak kunstig intelligens



Historien bak kunstig intelligens

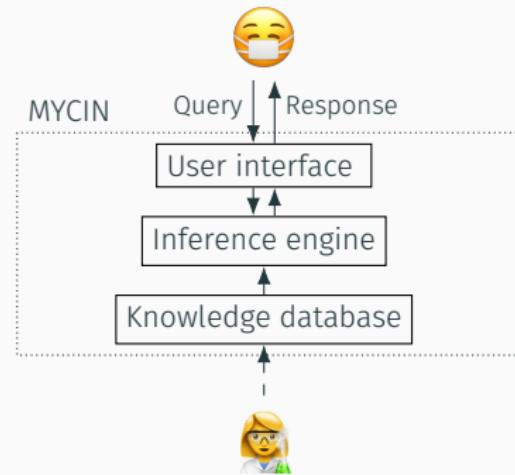
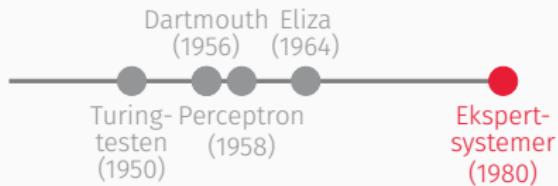


Talk to Eliza by typing your questions and answers in the input box.

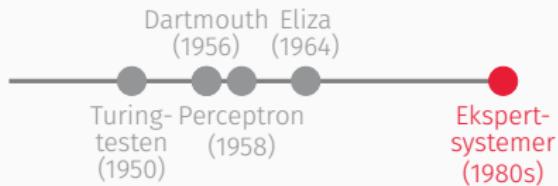
- * I'm feeling kind of sad today
- > Do you believe it is normal to be feeling kind of sad today?
- * No, I would prefer to be happy
- > Are you sure?
- * Yes
- > Are you sure about that?
- * Yes
- > Okay, now you're repeating yourself. That is not moving us forward.
- * You started
- > Oh... Let's move on to something else for a bit. started?

TYPE HERE

Historien bak kunstig intelligens

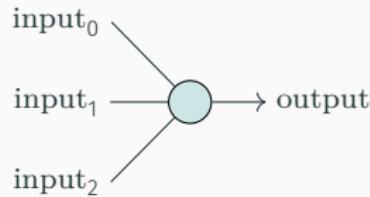
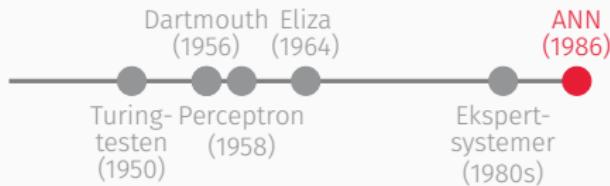


Historien bak kunstig intelligens

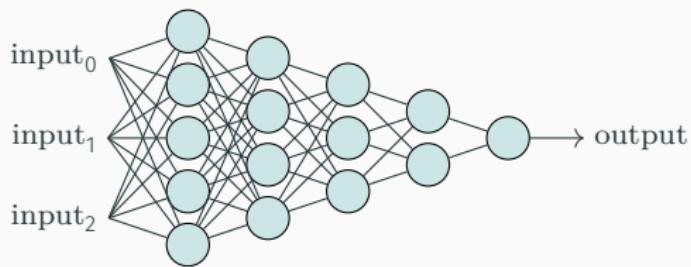
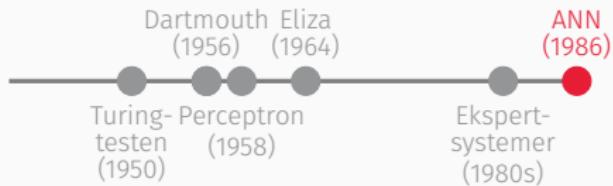


1) Patient's name: (first-last) ** FRED BRAUN	{background patient data}
2) Sex: ** M	
3) Age: ** 55	
4) Are there any cultures for Fred Braun which may be related to the present illness, and from which organisms have been grown successfully in the microbiology laboratory? ** Y	
-----CULTURE-1-----	
5) From what site was the specimen for CULTURE-1 taken? ** BLOOD	{typically identity is not yet known}
6) Please give the date and time when CULTURE-1 was obtained. (mo/ds/yr time) ** JUN 20, 1977	
The first organism isolated from the blood culture of 20-JUN-77 (CULTURE-1) will be referred to as:	
7) Enter the laboratory-reported identity of ORGANISM-1: ** UNKNOWN	
8) The stain (Gram or Ziehl-Neelsen acid-fast) of ORGANISM-1: ** NEG	
9) Is ORGANISM-1 a rod or coccus (etc.): ** ROD	
10) What is the form of the individual organisms (e.g. Lancet-shaped for cocci, fusiform for rods, etc.)? ** FUSIFORM	
{...more questions follow in order to gather sufficient information to infer the identity and significance of the infecting organisms...}	

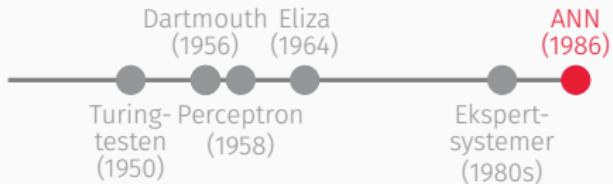
Historien bak kunstig intelligens



Historien bak kunstig intelligens



Historien bak kunstig intelligens



Learning representations by back-propagating errors

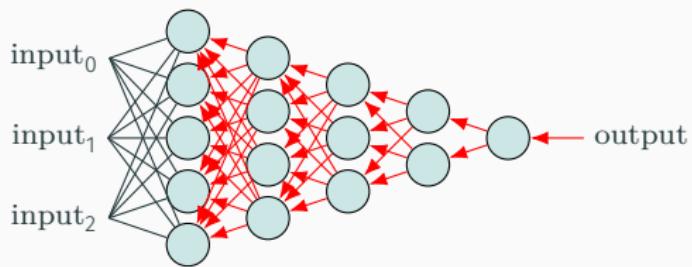
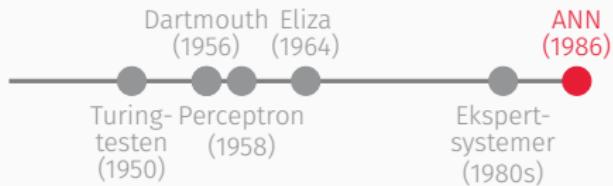
David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California,
San Diego, La Jolla, California 92093, USA

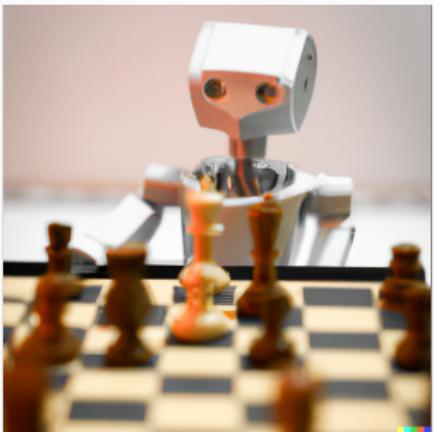
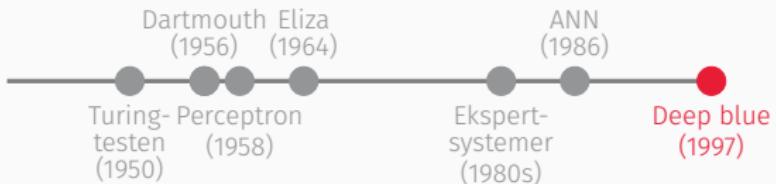
† Department of Computer Science, Carnegie-Mellon University,
Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure¹.

Historien bak kunstig intelligens



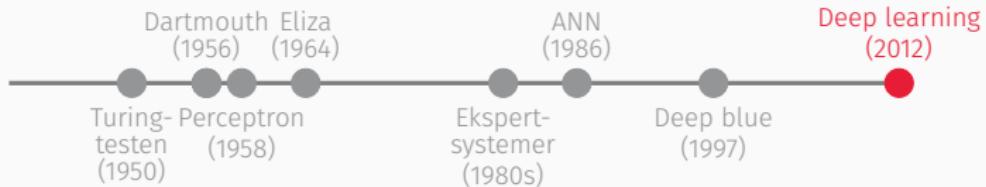
Historien bak kunstig intelligens



DALL-E: "A robot playing chess"

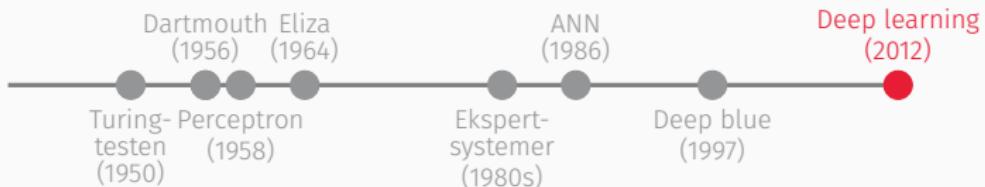
- IBMs Deep Blue slår Garry Kasparov og blir den første datamaskinen til å slå den sittende verdensmesteren i sjakk.
- Deep blue vant med 3½ poeng mot Kasparovs 2½ over 6 spill.
- Kasparov uttalte at "Deep Blue was intelligent the way your programmable alarm clock is intelligent."
- Kombinasjon av maskinlæring og hardkodet kunnskap fra sjakkeksperter.

Historien bak kunstig intelligens



Katt

Historien bak kunstig intelligens



Solsikke



Marihøne



Katt



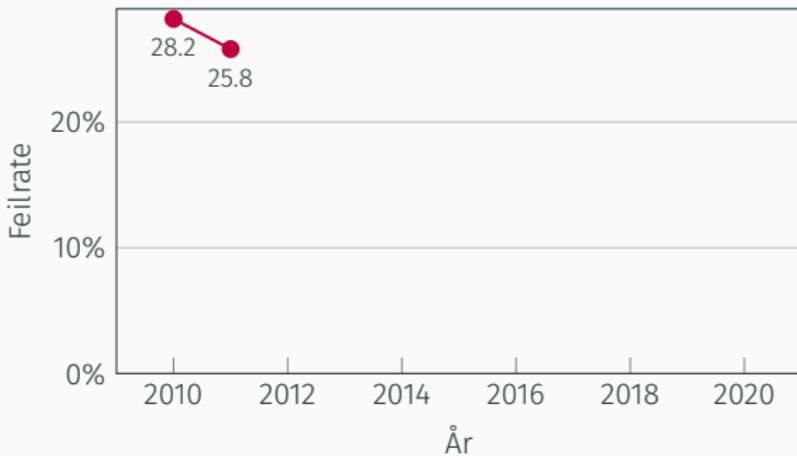
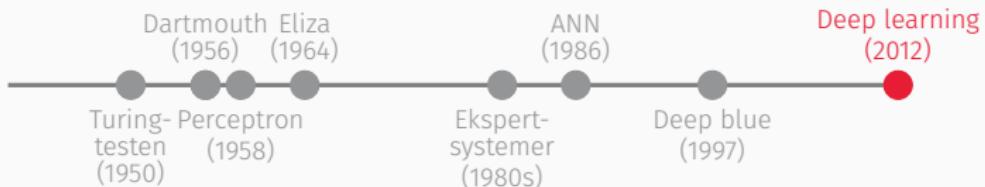
Fly



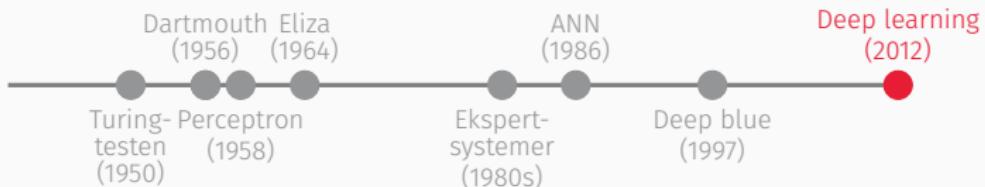
Hvithai

ImageNet: ~14m bilder, ~22k kategorier

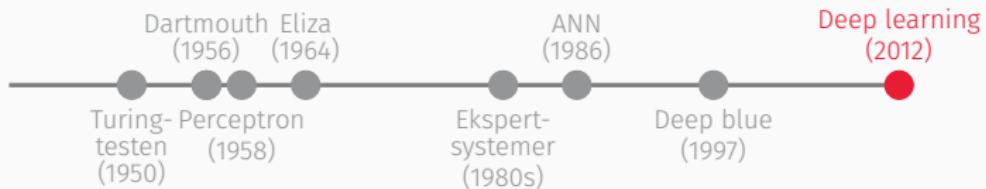
Historien bak kunstig intelligens



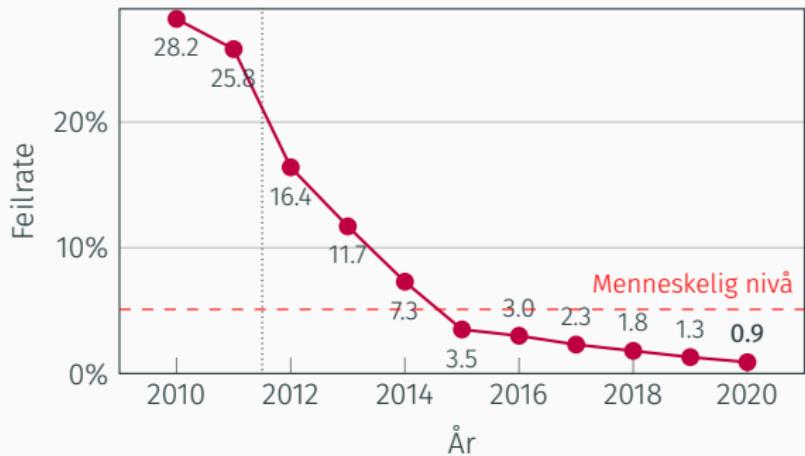
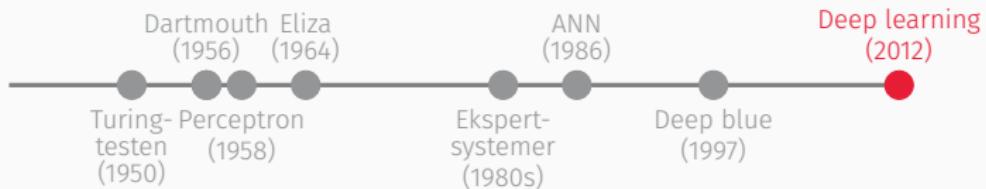
Historien bak kunstig intelligens



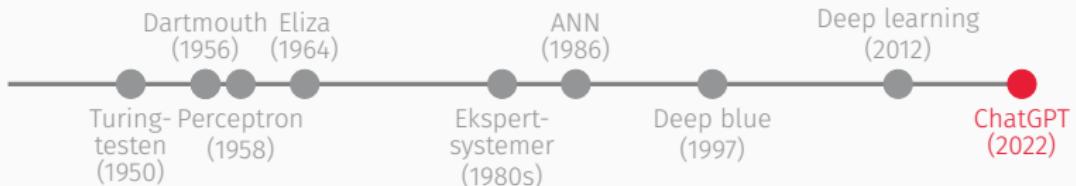
Historien bak kunstig intelligens



Historien bak kunstig intelligens



Historien bak kunstig intelligens



E

Kan du skrive et kjapt sammendrag av historien om kunstig intelligens til elever i andre klasse på vgs?



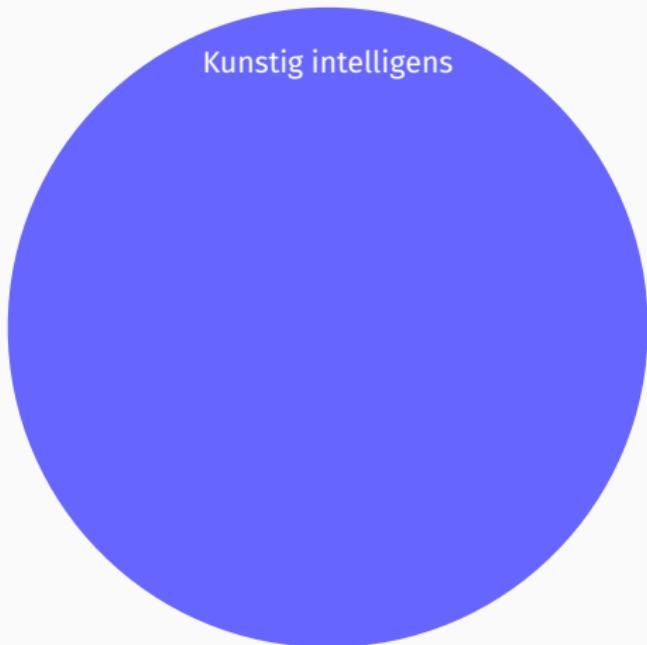
Selvfølgelig! Her er et enkelt og kort sammendrag av historien om kunstig intelligens (KI) egnet for elever i andre klasse på videregående skole:

Historien om Kunstig Intelligens (KI)

- Begynnelsen (1940-1950-tallet):** Alt startet da noen smarte folk tenkte: "Kan vi få maskiner til å tenke som mennesker?" De utviklet grunnleggende teorier for hvordan datamaskiner kunne løse problemer ved hjelp av matematikk.

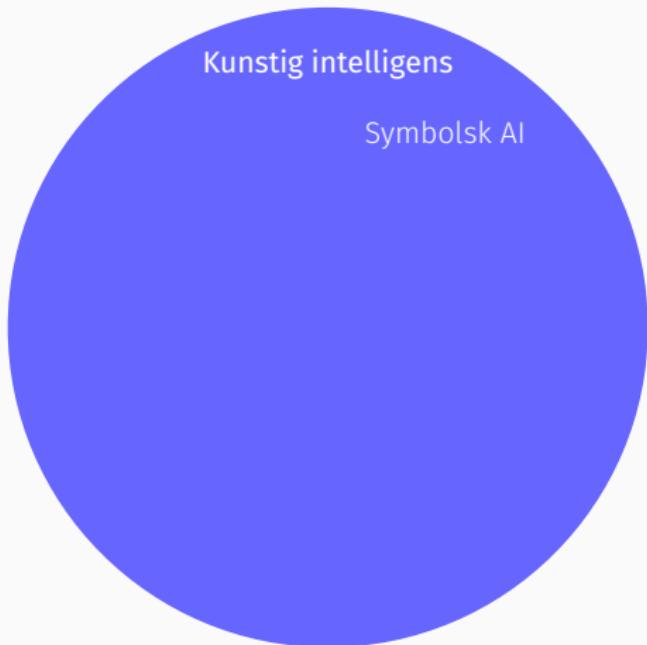
Terminologi og konsepter

Terminologi og konsepter: Begreper



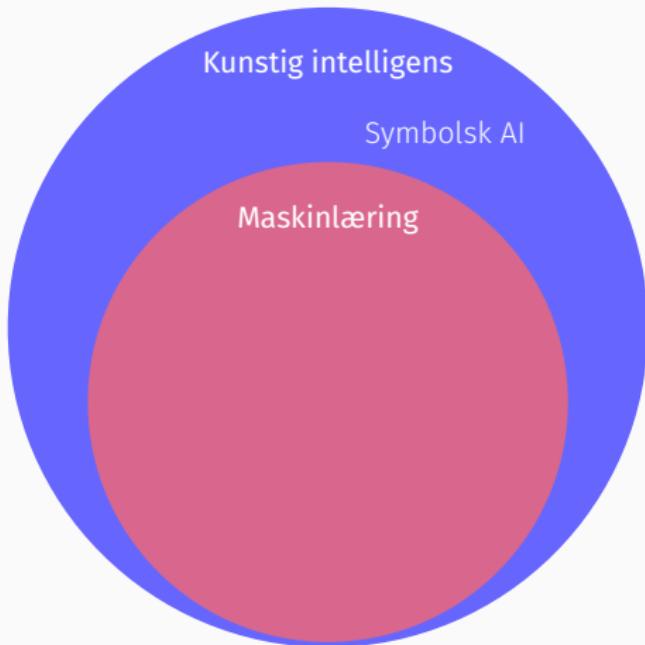
Kunstig intelligens:
Maskiner som løser oppgaver
som krever intelligens

Terminologi og konsepter: Begreper



Kunstig intelligens:
Maskiner som løser oppgaver
som krever intelligens

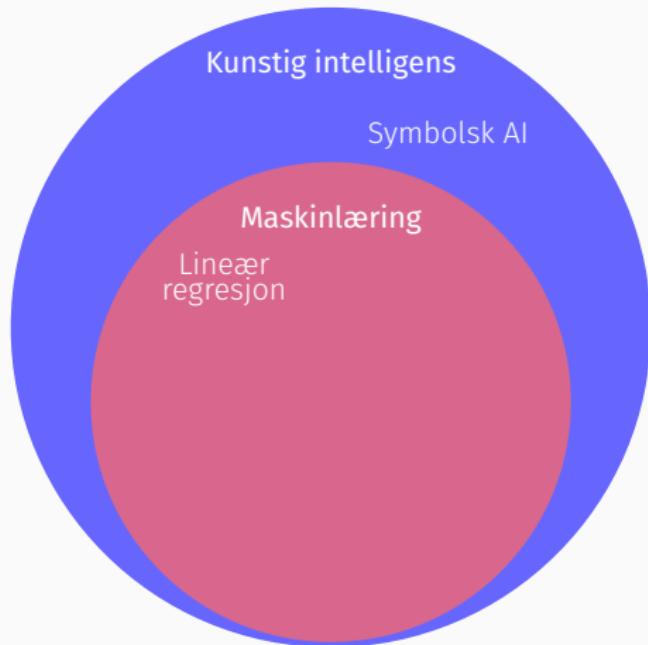
Terminologi og konsepter: Begreper



Kunstig intelligens:
Maskiner som løser oppgaver som krever intelligens

Maskinlæring:
Maskiner som løser problemer gjennom å lære seg å finne mønster i data

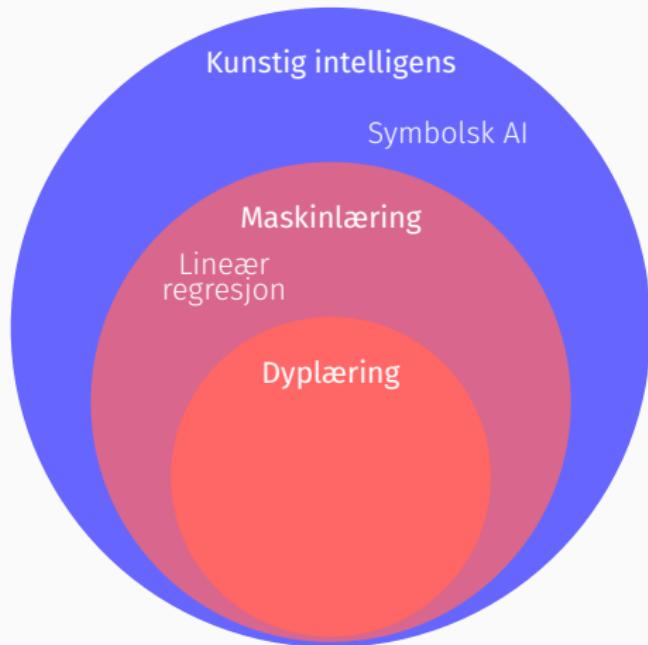
Terminologi og konsepter: Begreper



Kunstig intelligens:
Maskiner som løser oppgaver som krever intelligens

Maskinlæring:
Maskiner som løser problemer gjennom å lære seg å finne mønster i data

Terminologi og konsepter: Begreper

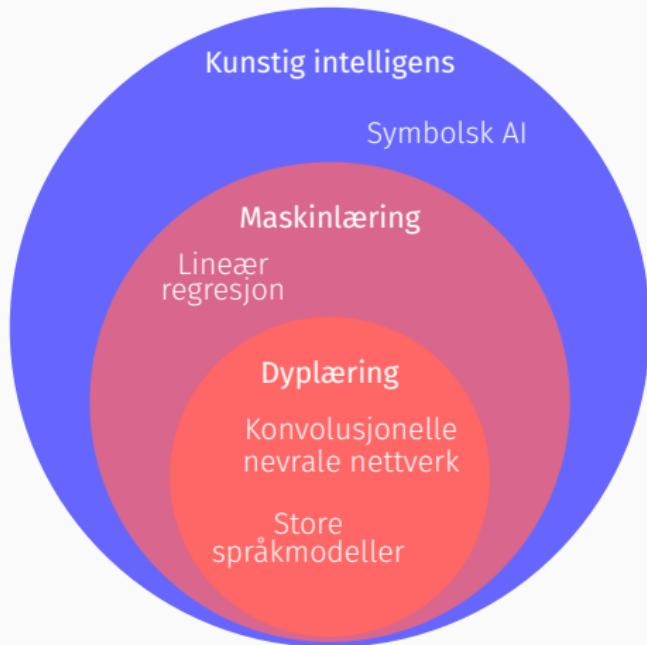


Kunstig intelligens:
Maskiner som løser oppgaver som krever intelligens

Maskinlæring:
Maskiner som løser problemer gjennom å lære seg å finne mønster i data

Dyplæring:
Maskinlæringsmodeller som er organisert i hierarkier
(≈ dype nevrale nettverk)
inspirert av hjernen

Terminologi og konsepter: Begreper



Kunstig intelligens:

Maskiner som løser oppgaver som krever intelligens

Maskinlæring:

Maskiner som løser problemer gjennom å lære seg å finne mønster i data

Dyplæring:

Maskinlæringsmodeller som er organisert i hierarkier
(≈ dype nevrale nettverk)
inspirert av hjernen

Konvolusjonelle nevrale nettverk:

Nevrale nettverk for bildedata

Store språkmodeller:

Nevrale nettverk for språkdata (ChatGPT)

Terminologi og konsepter: Veiledning

Veiledet læring
(supervised learning)



→ Katt



→ Hund



→ Katt



→ Hund

Ikke-veiledet læring
(unsupervised learning)



Terminology: Smal og generell AI

Smal (svak)

Generell (sterk)

I stand til å løse et bredere spekter med problemer innenfor flere domener



Mer spesifikk

Mer generell

Terminology: Smal og generell AI

Smal (svak)

Generell (sterk)

I stand til å løse et bredere spekter med
problemer innenfor flere domener



Mer spesifik



Mer generell



Terminology: Smal og generell AI

Smal (svak)

Generell (sterk)

I stand til å løse et bredere spekter med
problemer innenfor flere domener



Bilde-
diagnostikk

Forsikrings-
prising

Dokument-
lesing



Mer spesifikk



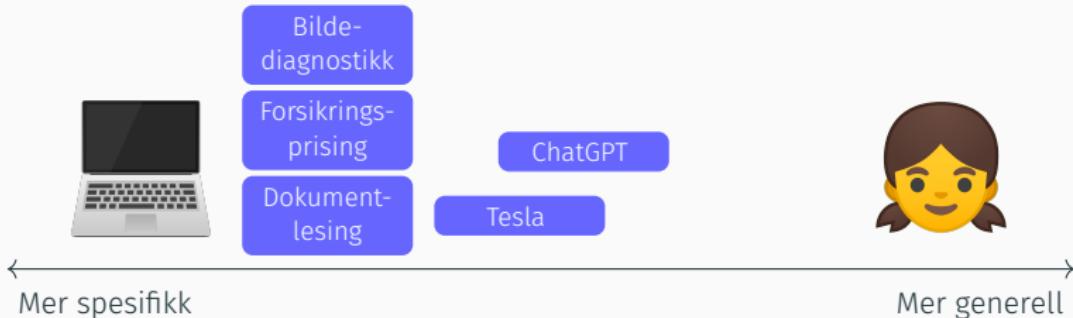
Mer generell

Terminology: Smal og generell AI

Smal (svak)

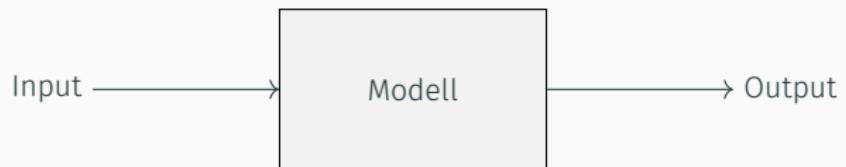
Generell (sterk)

I stand til å løse et bredere spekter med
problemer innenfor flere domener

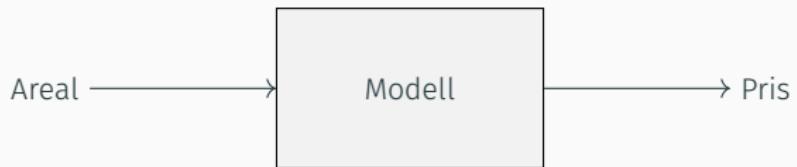


Teori

Teori: Maskinlæringsmodeller



Teori: Maskinlæringsmodeller



Teori: Maskinlæringsmodeller

Finn - Møblermarked

Søkresultat

- Vårsted og Tømmer (5)
- Verftsted (3)
- Villave (2)

Søkstatus

- Til salgs (310)
- Søkt siden 3 dager (9)
- Kommer fra salg (1)

Nyhetstidslinje

- Brukt bobil (12)
- Nybils (237)

Prisnivå

- Høyre
- Til venstre

Søk

Oppdatert

- Høyre
- Til venstre

Søk

Forsidepriser per måned

- Høyre
- Til venstre

Søk

Stereotyper

- Høyre
- Til venstre

Søk

Antall soverom

- 0+1
- 1+
- 2+
- 3+
- 4+
- 5+

Søk

Baigtekjøkken

- Høyre
- Til venstre

Søk

Baigetyper

- Leilighet (544)
- Garage/Parkeringsplass (11)
- Rekkehus (2)
- Endebolig (1)
- Tverrleilighet (2)
- Lager/Logistik (2)
- Andre (2)

Være

Oppdatert

- Akjøp (5)
- Andel (124)
- Gav (54) (søkstart) (233)

Varetyper

Lillestrømsbygning Løren

Løren - Tilataende 3-roms med god planleining og flott beliggenhet - Solrik...
73 m² • 6 400 000 kr
Tilgang: 4 til 12 døgn • Ballosgang: 4 til 12 døgn • Van: Detaljer • Logget: - 2 sekvenser

Visning: 32 ganger (12/2021)

SP Grønne & Parkens Grünerløkka

Rodeløkka / Grünerløkka - Lys, luftig 3-roms hjemmetil med innre gjerd i...
67 m² • 5 500 000 kr
Tilgang: 3 til 12 døgn • Ballosgang: 4 til 12 døgn • Van: Detaljer • Logget: - 2 sekvenser

Visning: 22 ganger (12/2021)

Hansens gate 6, Oslo

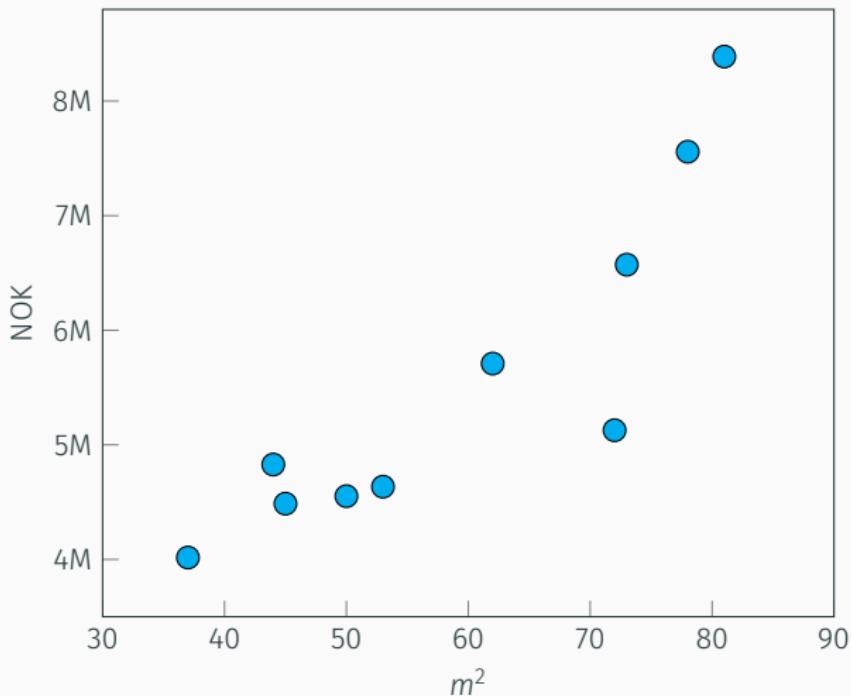
Grønnelekka - Lys 2-roms med stort potensial! Nordvest balkong - Heis...
46 m² • 4 100 000 kr
Tilgang: 4 til 12 døgn • Ballosgang: 3 til 12 døgn • Van: Detaljer • Logget: - 2 sekvenser

Visning: 22 ganger (12/2021)

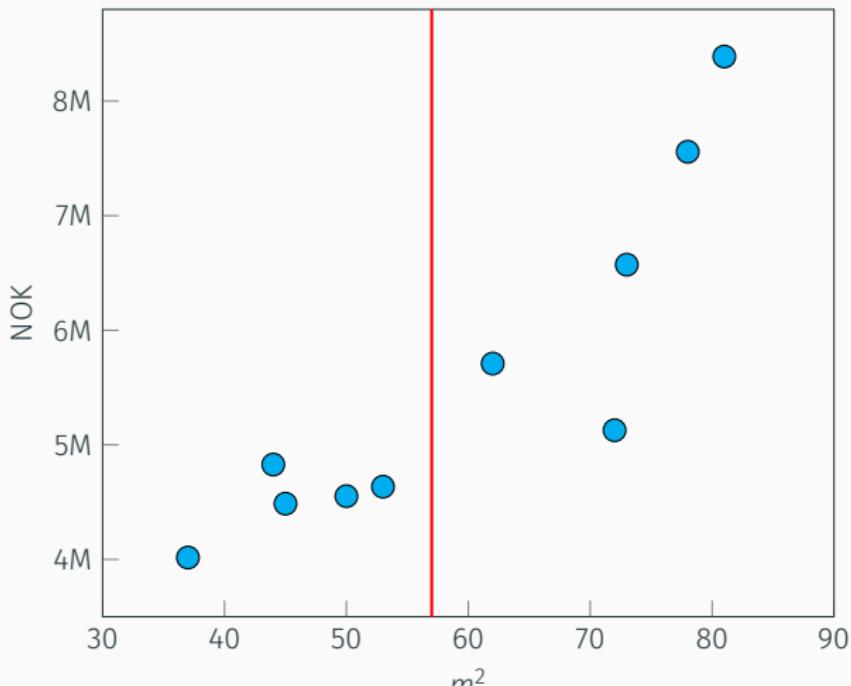
Teori: Maskinlæringsmodeller

m^2	Pris
72	5.127.379
50	4.552.170
45	4.486.654
62	5.709.276
53	4.634.912
81	8.388.570
44	4.828.170
78	7.557.770
37	4.016.520
73	6.572.351

Teori: Maskinlæringsmodeller

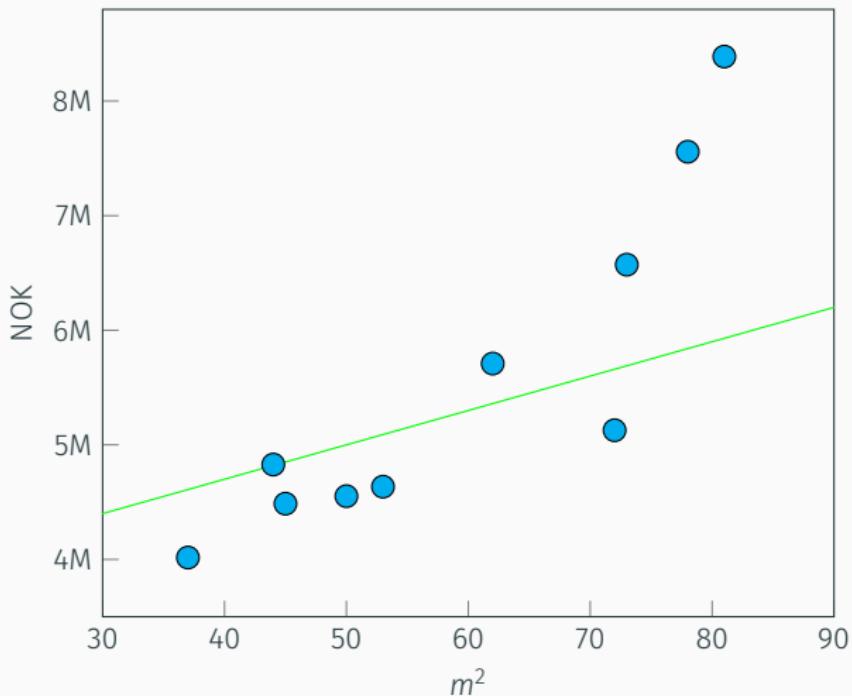


Teori: Maskinlæringsmodeller

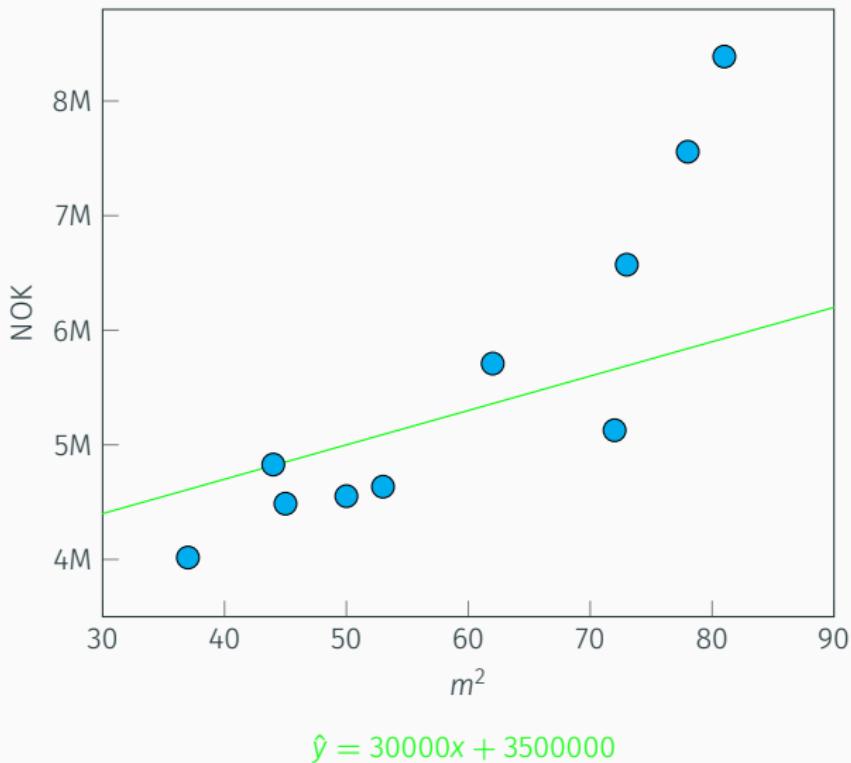


$$\hat{y} = f(57)$$

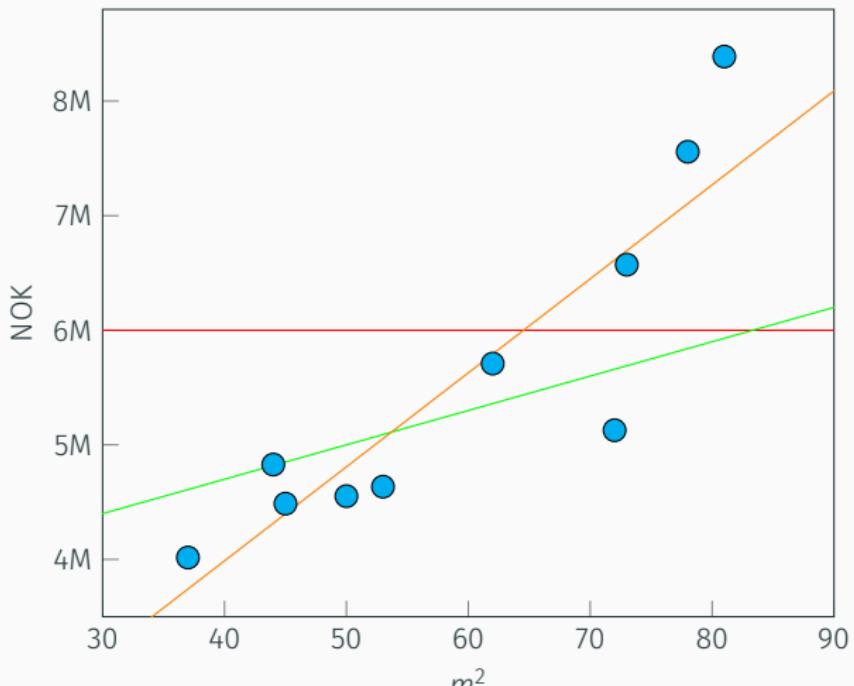
Teori: Maskinlæringsmodeller



Teori: Maskinlæringsmodeller



Teori: Maskinlæringsmodeller

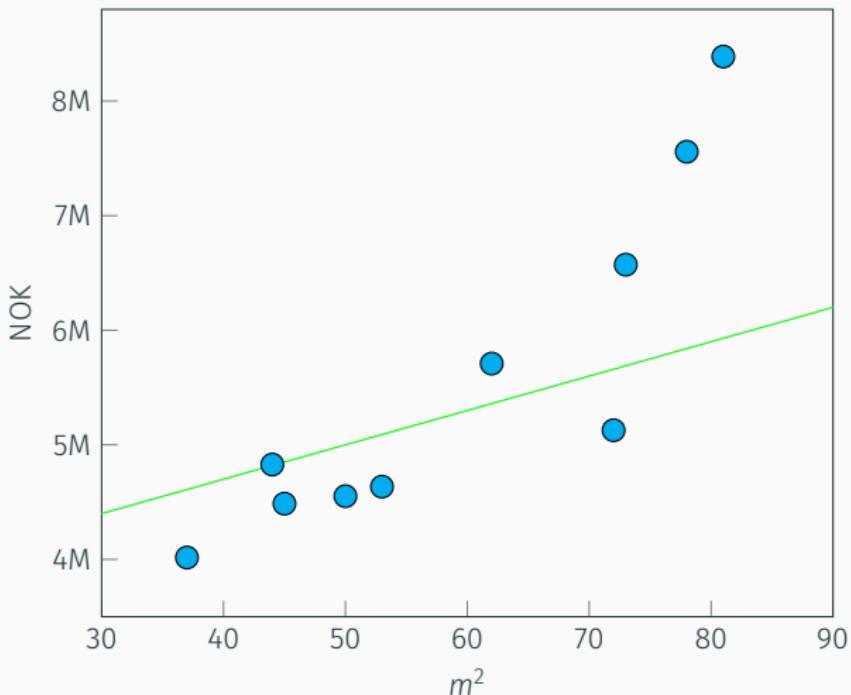


$$\hat{y} = 0x + 6000000$$

$$\hat{y} = 30000x + 3500000$$

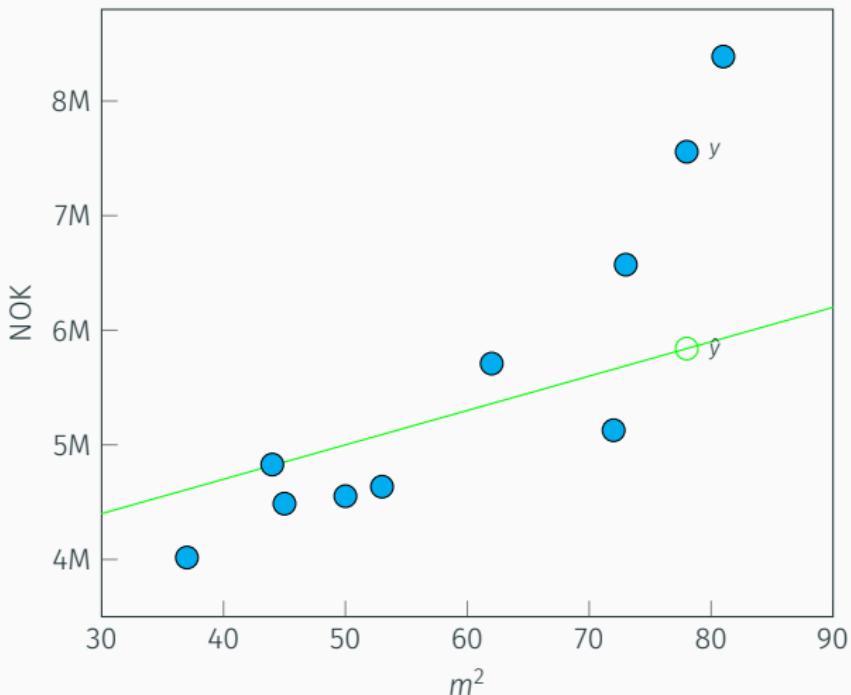
$$\hat{y} = 82031x + 706495$$

Teori: Kostfunksjon



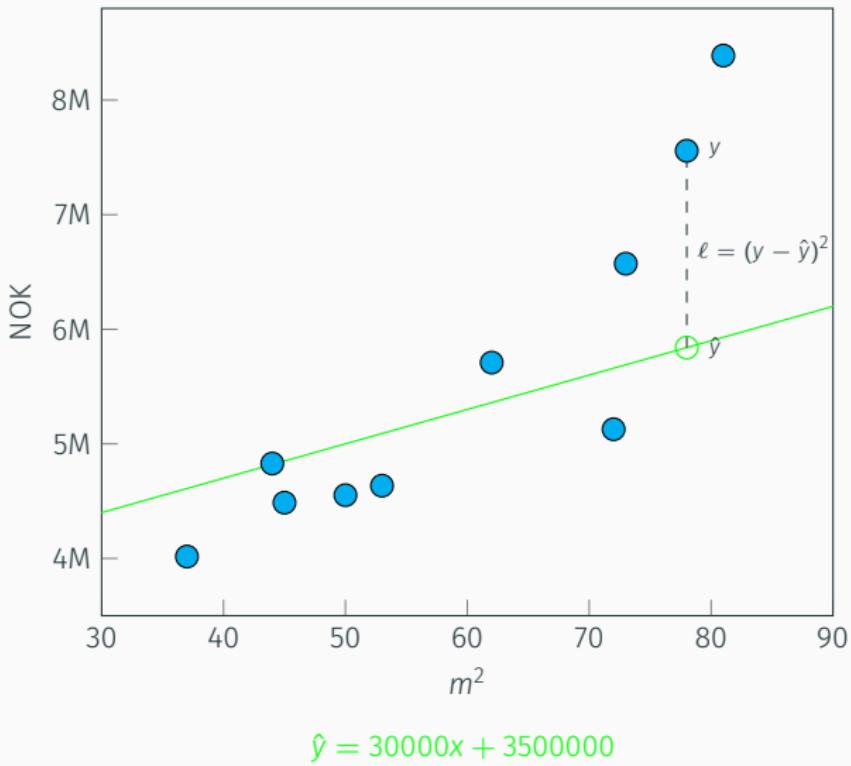
$$\hat{y} = 30000x + 3500000$$

Teori: Kostfunksjon

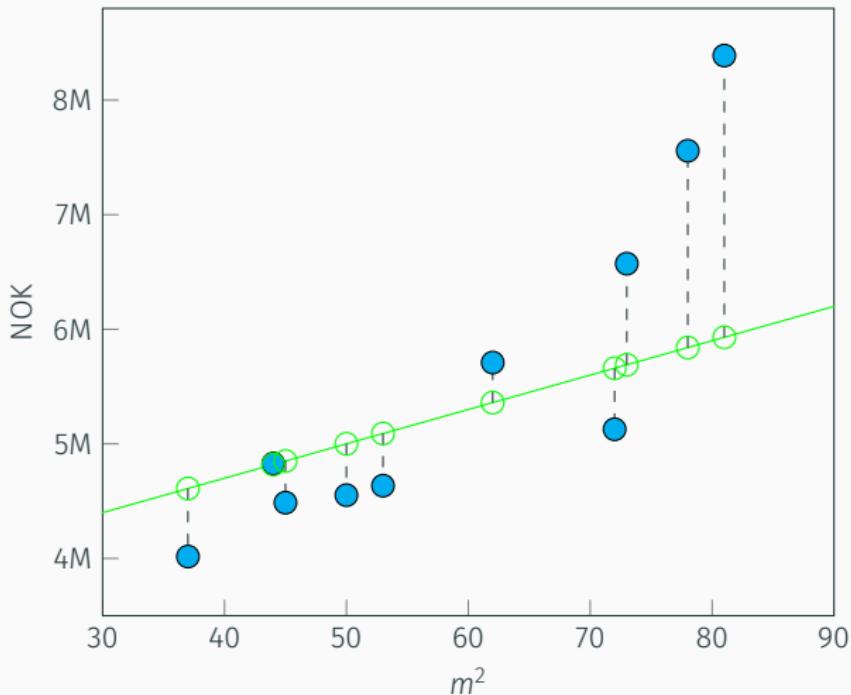


$$\hat{y} = 30000x + 3500000$$

Teori: Kostfunksjon



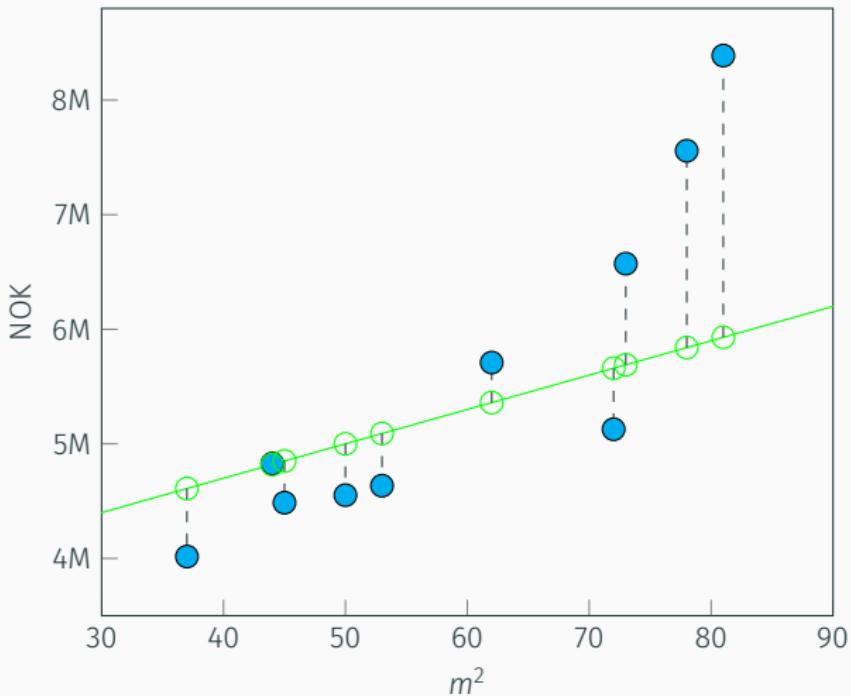
Teori: Kostfunksjon



$$\hat{y} = 30000x + 3500000$$

$$\ell = \sum(y - \hat{y})^2$$

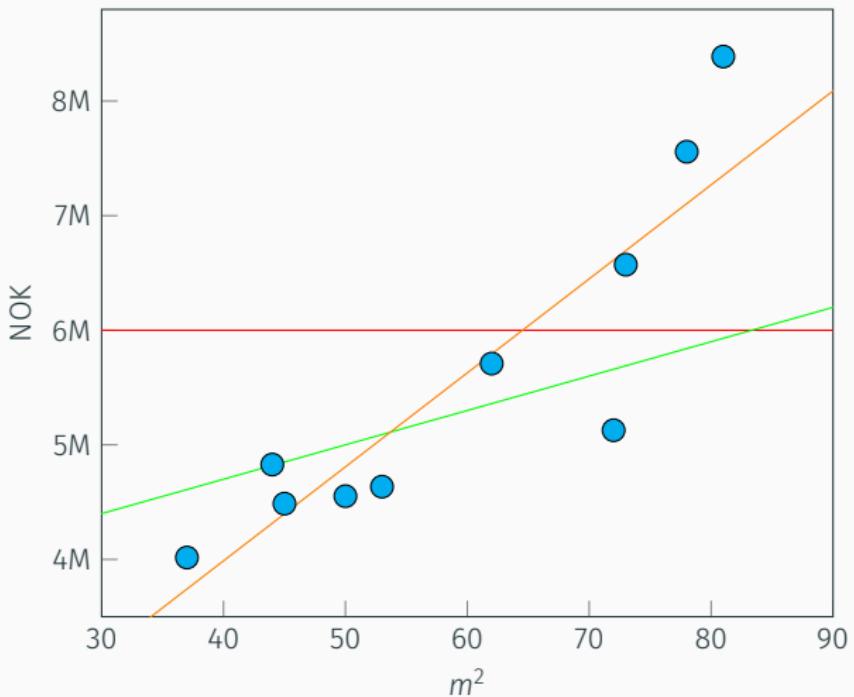
Teori: Kostfunksjon



$$\hat{y} = 30000x + 3500000$$

$$\ell = 1.10 \times 10^{13}$$

Teori: Kostfunksjon

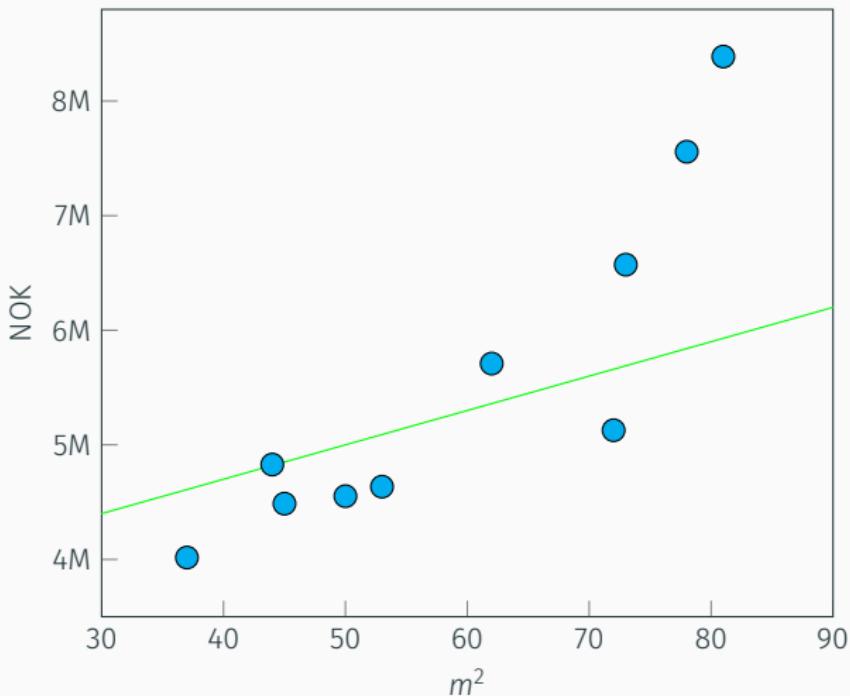


$$\hat{y} = 0x + 6000000$$
$$\ell = 2.08 \times 10^{13}$$

$$\hat{y} = 30000x + 3500000$$
$$\ell = 1.10 \times 10^{13}$$

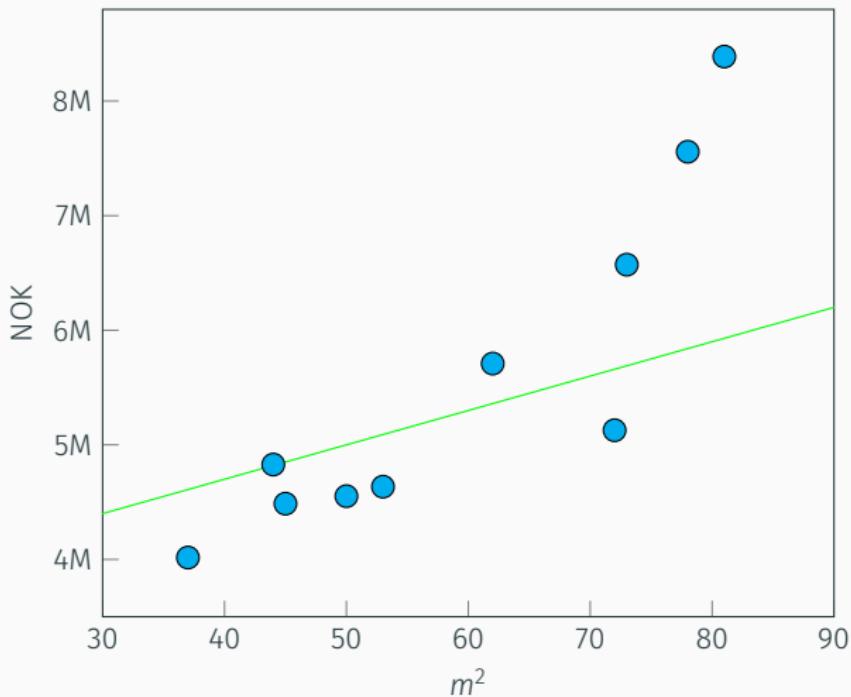
$$\hat{y} = 82031x + 706495$$
$$\ell = 4.09 \times 10^{12}$$

Teori: Trening



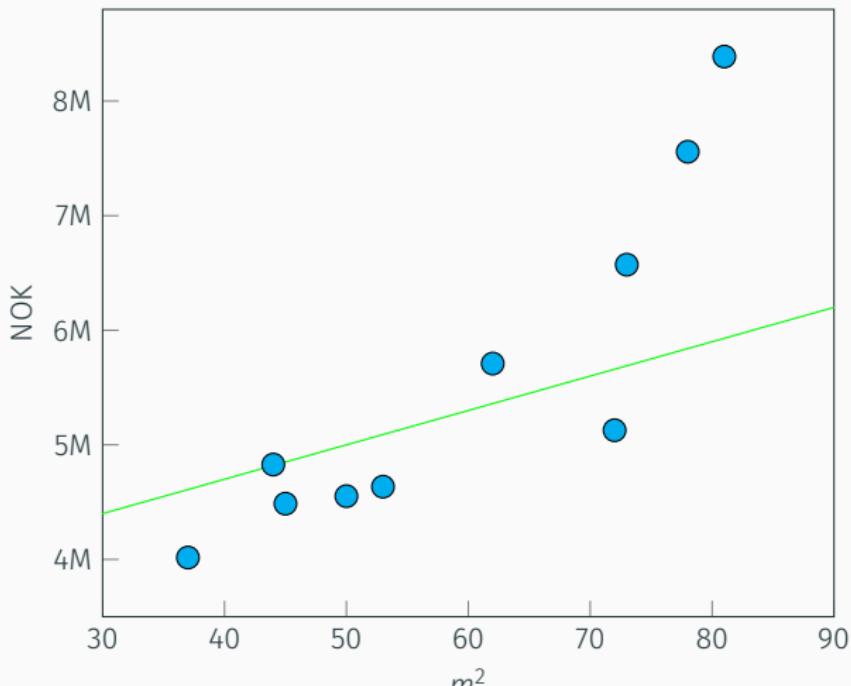
$$\hat{y} = wx + b$$
$$\ell = \sum(y - \hat{y})^2$$

Teori: Trening



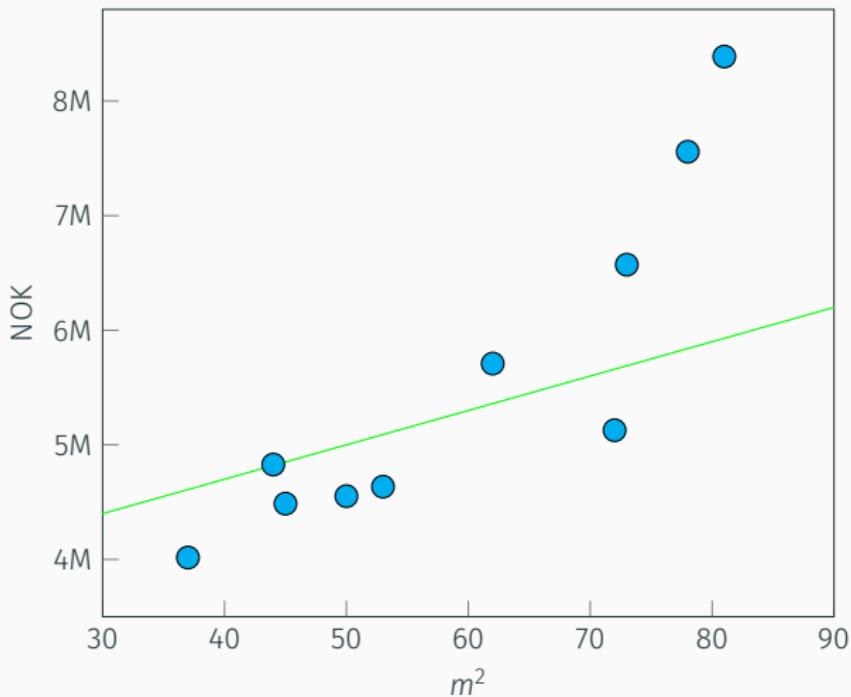
$$\hat{y} = wx + b$$
$$\ell = \sum(y - \hat{y})^2$$

Teori: Trening



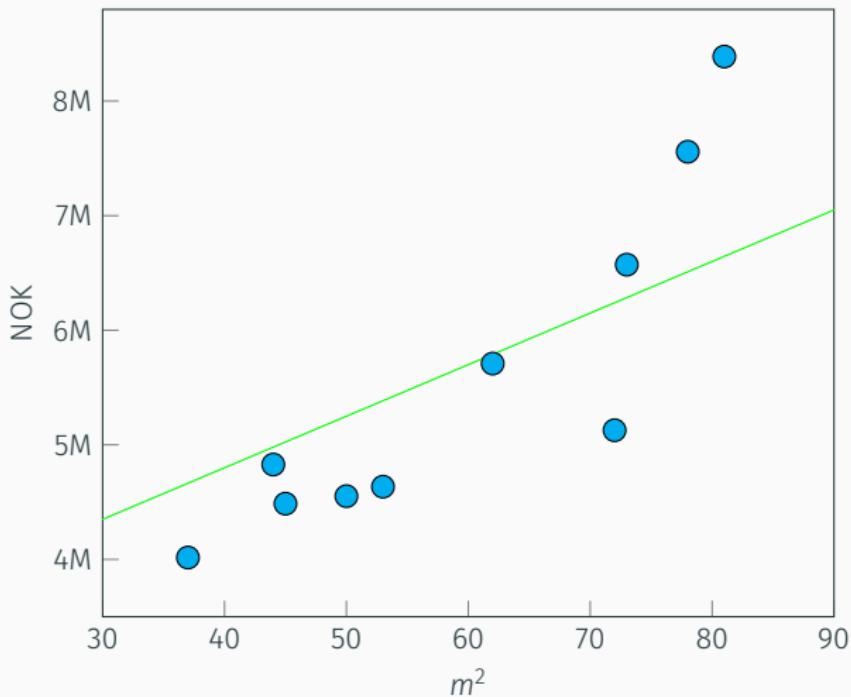
$$\ell = \sum(y - (wx + b))^2$$

Teori: Trening



$$1.10 \times 10^{13} = \sum(y - (30000x + 3500000))^2$$

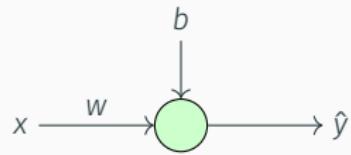
Teori: Trening



$$7.24 \times 10^{12} = \sum(y - (45000x + 3000000))^2$$

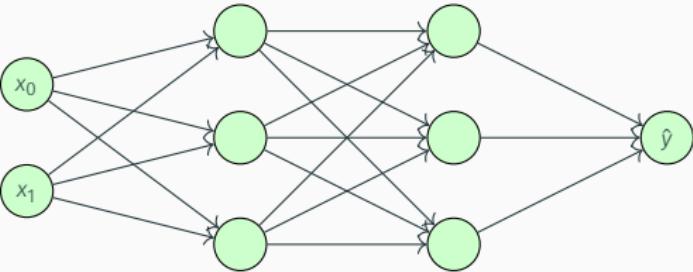
Demo!

Teori: Nevrale nettverk



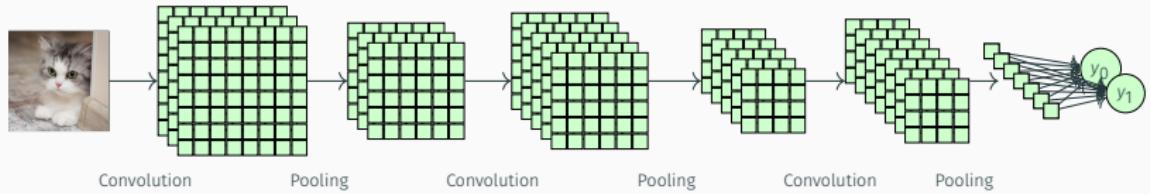
$$\hat{y} = wx + b$$

Teori: Nevrale nettverk

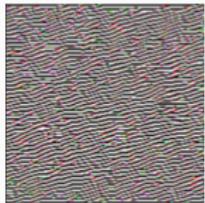
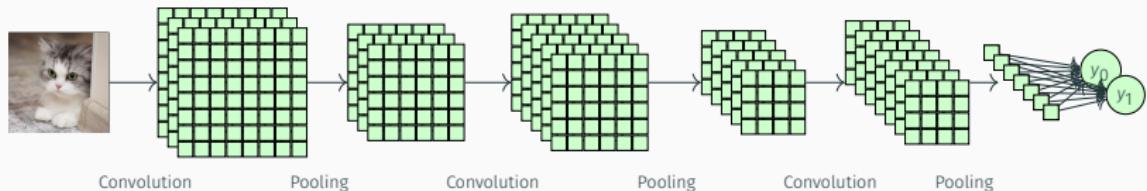


$$\begin{aligned}\hat{y} = & \max(0, w_{0,0}^2 * \max(0, w_{0,0}^1 * \max(0, w_{0,0}^0 * x_0 + w_{1,0}^0 * x_1 + b_{0,0}) + \\ & w_{1,0}^1 * \max(0, w_{0,1}^0 * x_0 + w_{1,1}^0 * x_1 + b_{0,1}) + \\ & w_{2,0}^1 * \max(0, w_{0,2}^0 * x_0 + w_{1,2}^0 * x_1 + b_{0,2}) + \\ & b_{1,0}) + \\ & w_{1,0}^2 * \max(0, w_{0,1}^1 * \max(0, w_{0,0}^0 * x_0 + w_{1,0}^0 * x_1 + b_{0,0}) + \\ & w_{1,1}^1 * \max(0, w_{0,1}^0 * x_0 + w_{1,1}^0 * x_1 + b_{0,1}) + \\ & w_{2,1}^1 * \max(0, w_{0,2}^0 * x_0 + w_{1,2}^0 * x_1 + b_{0,2}) + \\ & b_{1,1}) + \\ & w_{2,0}^2 * \max(0, w_{0,2}^1 * \max(0, w_{0,0}^0 * x_0 + w_{1,0}^0 * x_1 + b_{0,0}) + \\ & w_{1,2}^1 * \max(0, w_{0,1}^0 * x_0 + w_{1,1}^0 * x_1 + b_{0,1}) + \\ & w_{2,2}^1 * \max(0, w_{0,2}^0 * x_0 + w_{1,2}^0 * x_1 + b_{0,2}) + \\ & b_{1,2}) + \\ & b_2)\end{aligned}$$

Teori: Konvolusjonelle nevrale nettverk

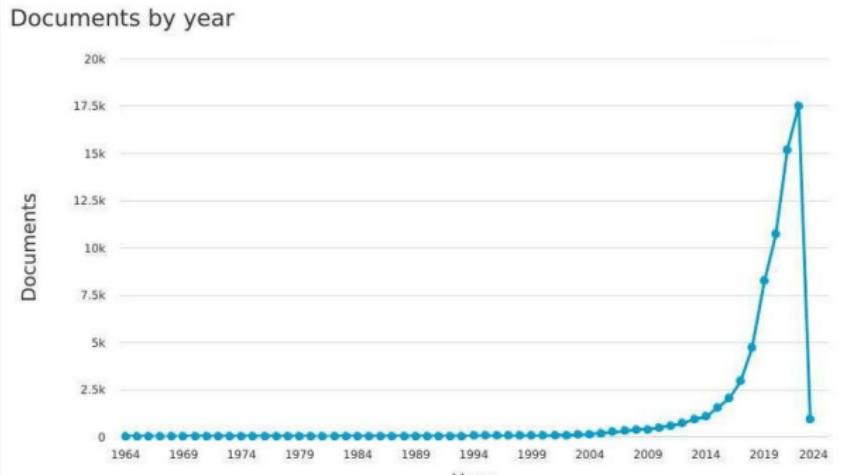


Teori: Konvolusjonelle nevrale nettverk



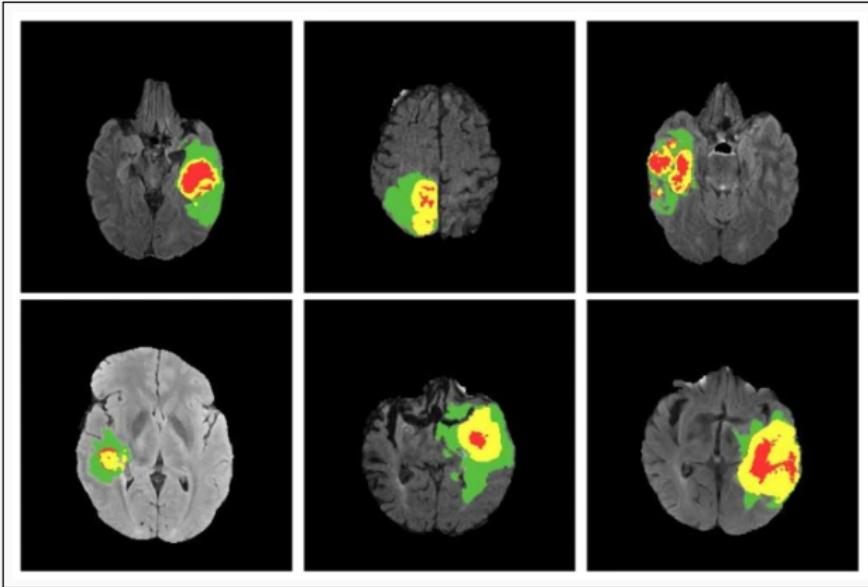
AI i hjernehorskning

AI i hjernehforskning: Oversikt



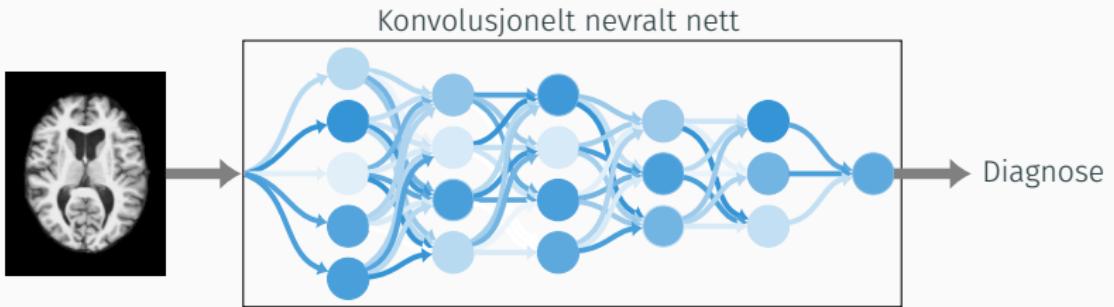
Artikler om dyplæring i medisin per år

AI i hjernehjerneforskning: Segmentering

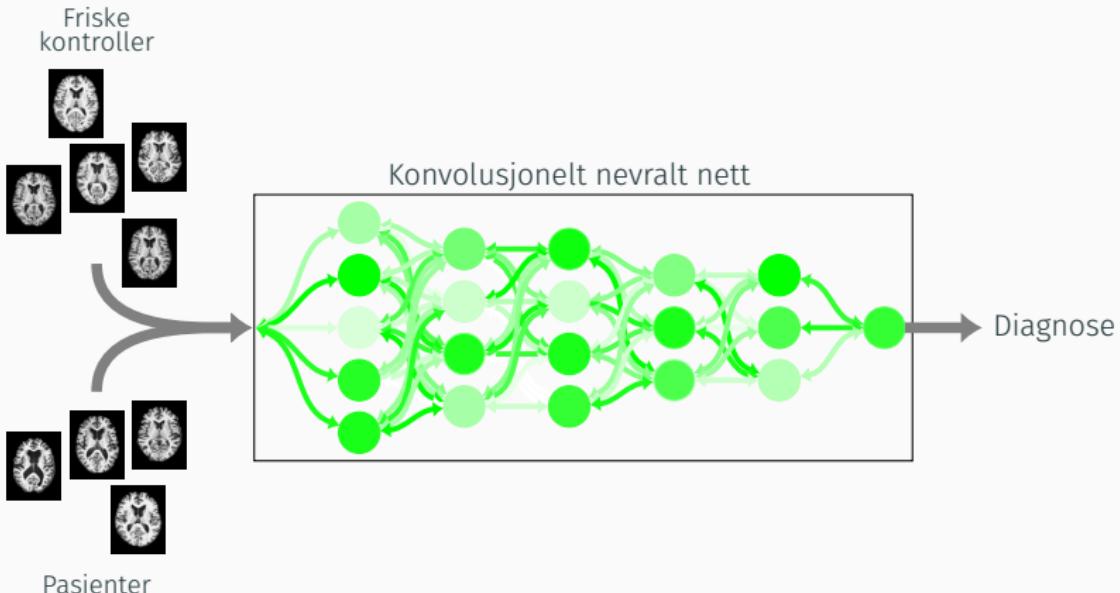


Segmentering av hjernesvulster (glioblastoma) basert på MR-bilder

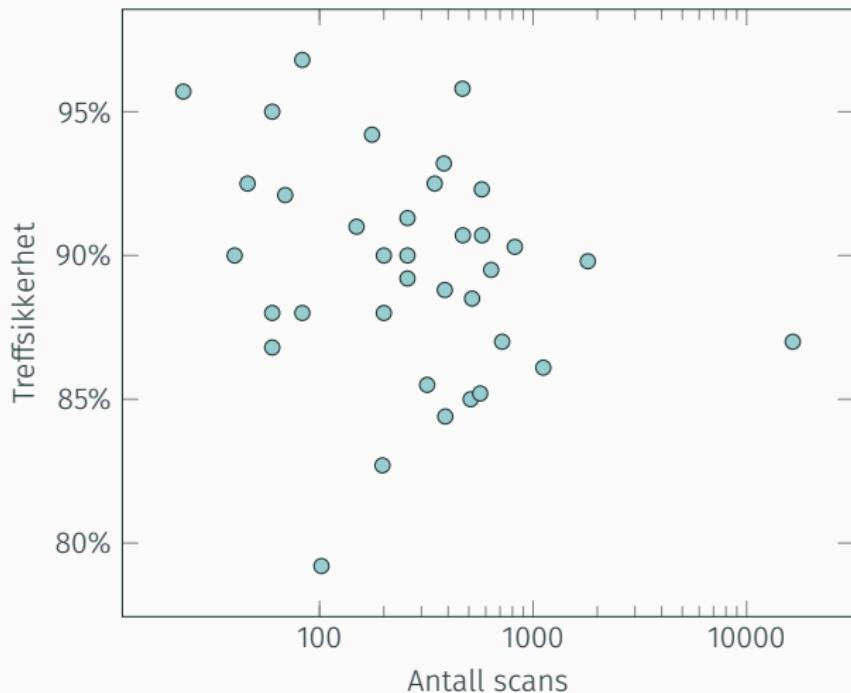
AI i hjernehjerneforskning: Diagnostisk klassifisering



AI i hjernehjerneforskning: Diagnostisk klassifisering

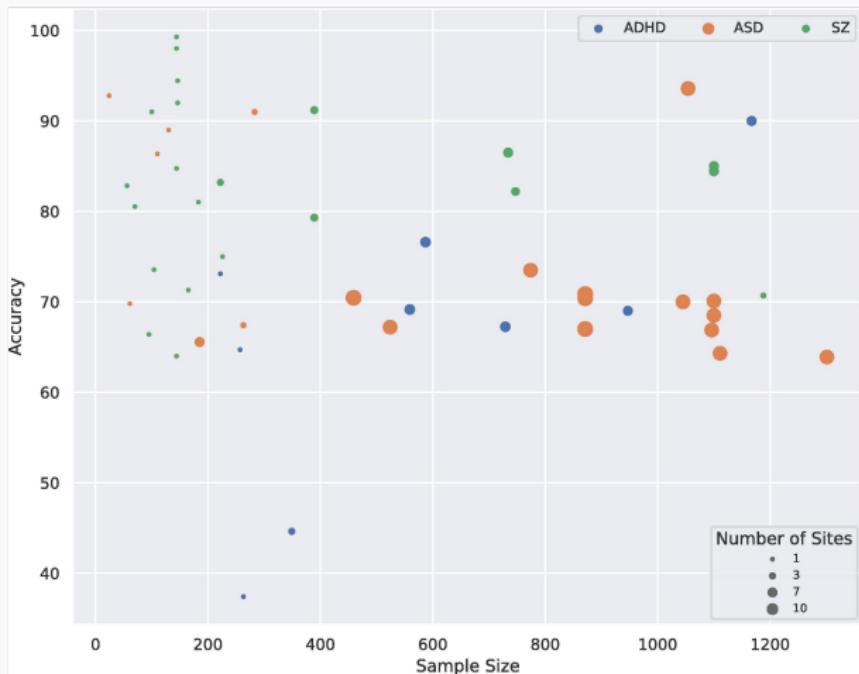


AI i hjernehjerneforskning: Diagnostisk klassifisering



Klassifisering av Alzheimer's sykdom

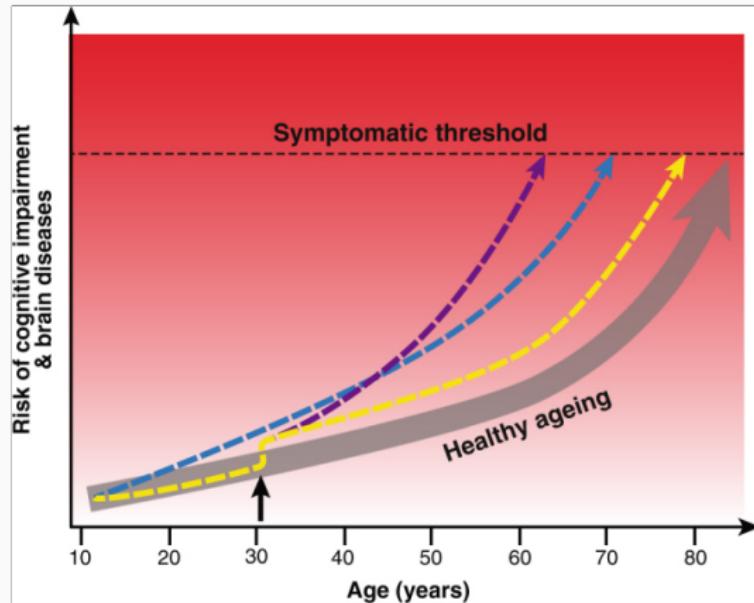
AI i hjernehorskning: Diagnostisk klassifisering



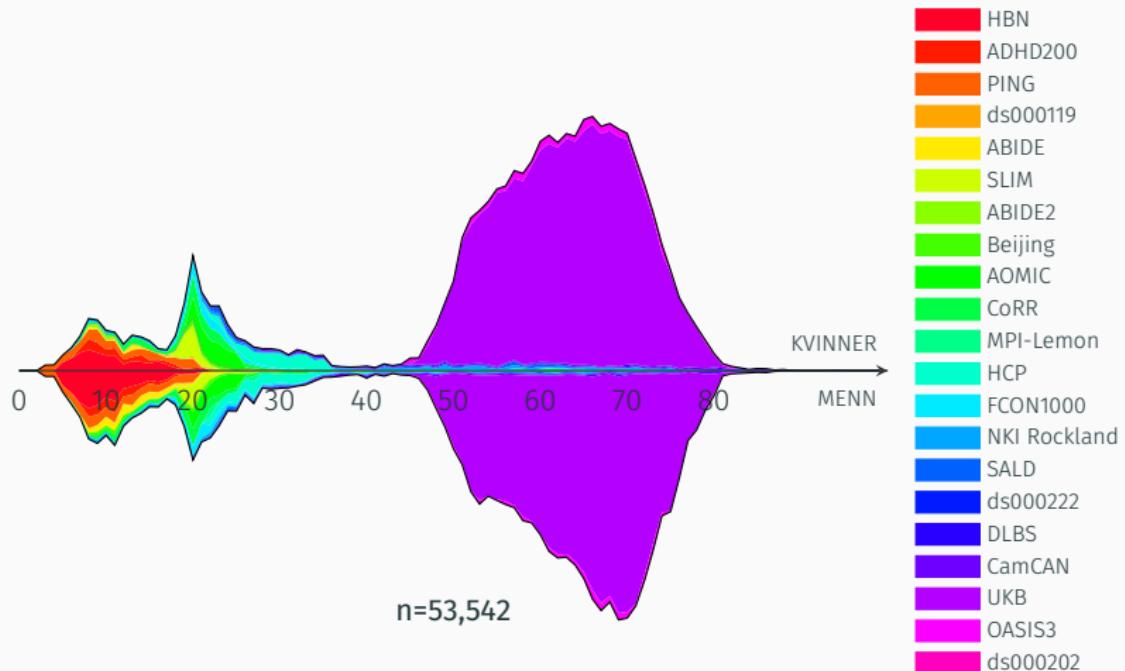
Hvorfor er diagnostisk klassifisering i mange tilfeller lite hensiktsmessig?

- Forutsetter at pasientene allerede er syke.
- Forutsetter at vi allerede vet at pasientene er syke.
- Forutsetter at det finnes en vel-definert pasient-gruppe.
- Vi har en god klassifikator. Hva nå?

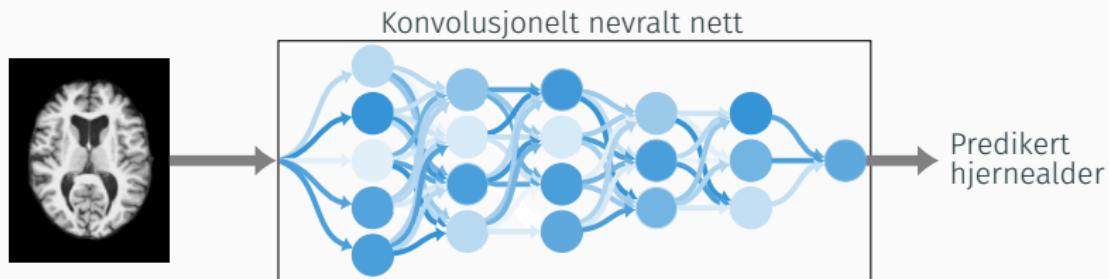
AI i hjernehørsel: Hjernealder



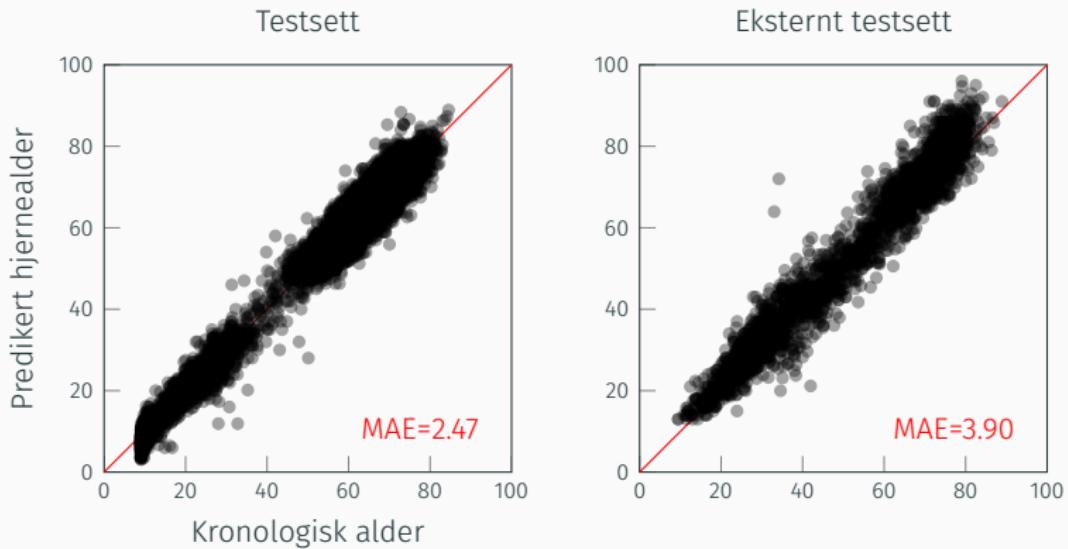
AI i hjernehforskning: Hjernealder



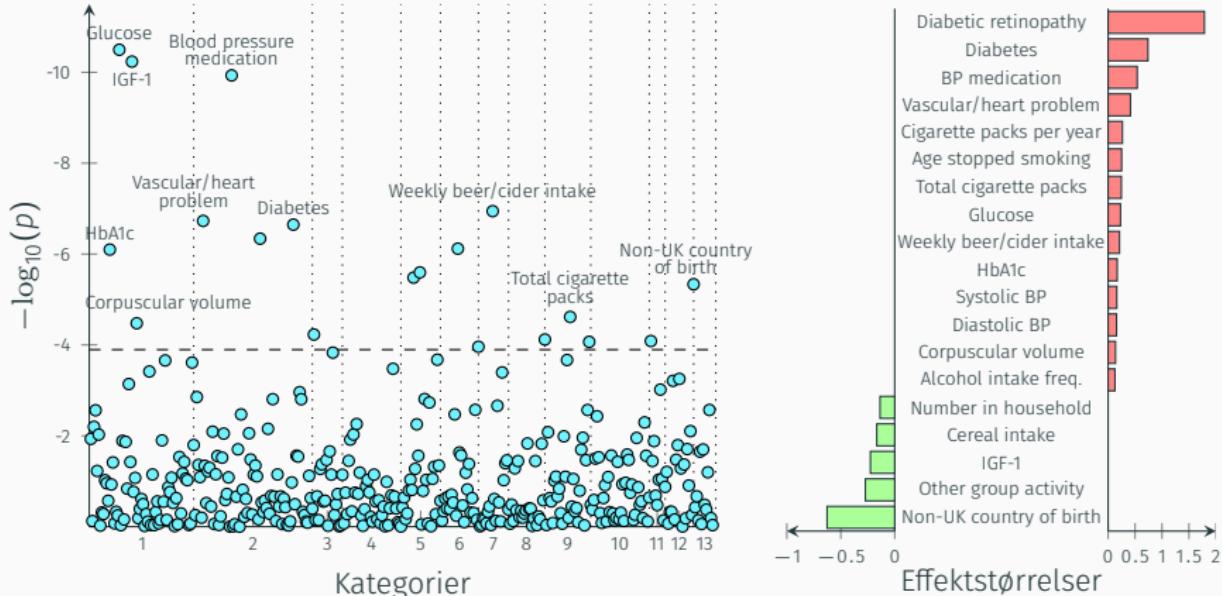
AI i hjernehforskning: Hjernealder



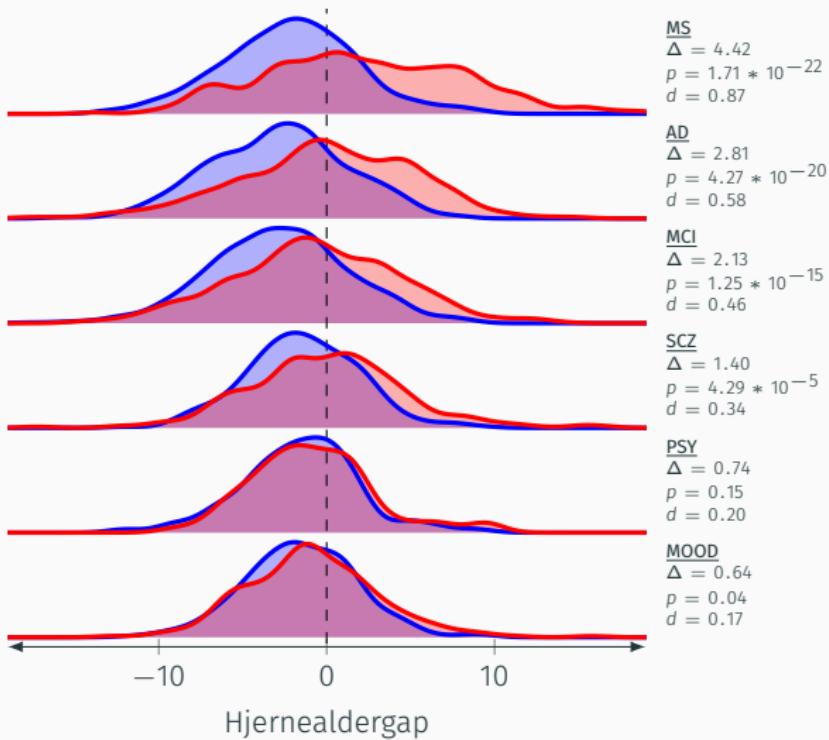
AI i hjernehforskning: Hjernealder



AI i hjernehalskning: Hjernealder



AI i hjernehforskning: Hjernealder

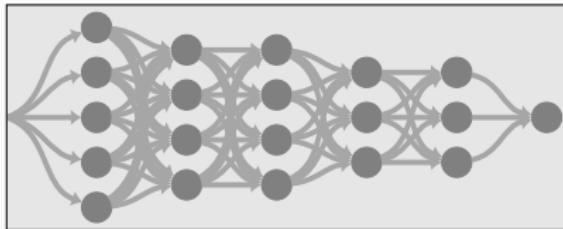


Hjernealder er et potensielt data-derivert mål på hjernens generelle helse

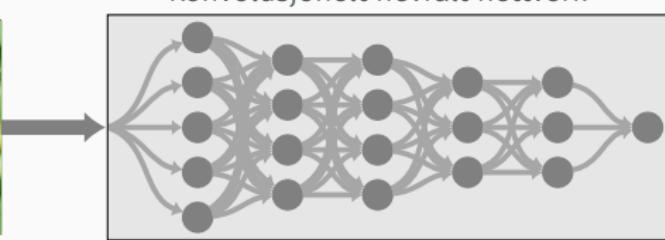
- + Oppsummerer all informasjonen i en hjernescan i ett tall.
- + Intuitivt.
- + Samvarierer med andre faktorer (hjerteproblemer, alkoholinntak, røyking).
- + Varierer mellom sykdomsgrupper (multipel sklerose, Alzheimer's sykdom).
- Ikke et direkte mål på helsetilstand.
- Usikkert hvordan det skal tolkes.
- Hva er det modellen ser på?

AI i hjernehjerneforskning: Forklarbar AI

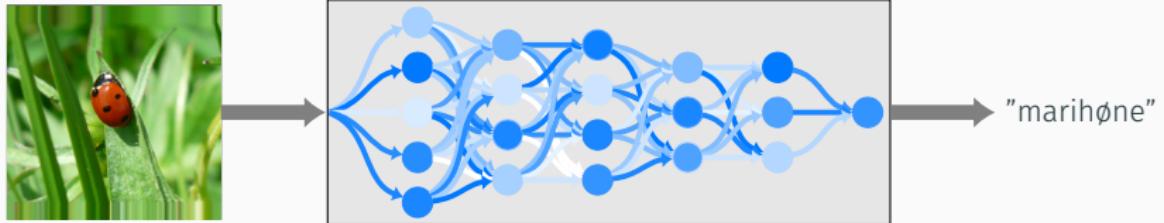
Konvolusjonelt nevralgt nettverk

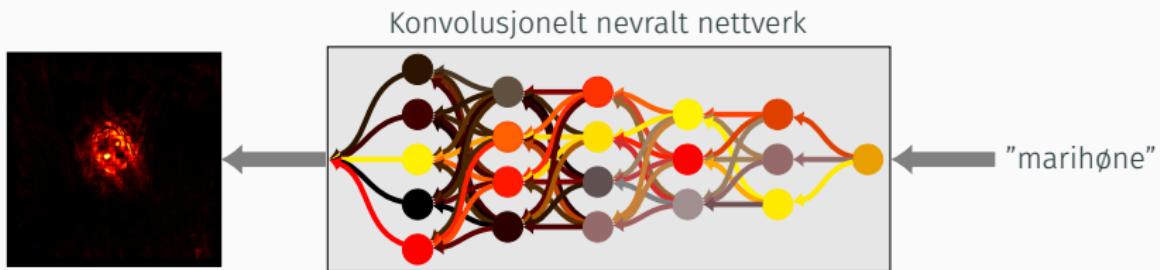


AI i hjernehforskning: Forklarbar AI

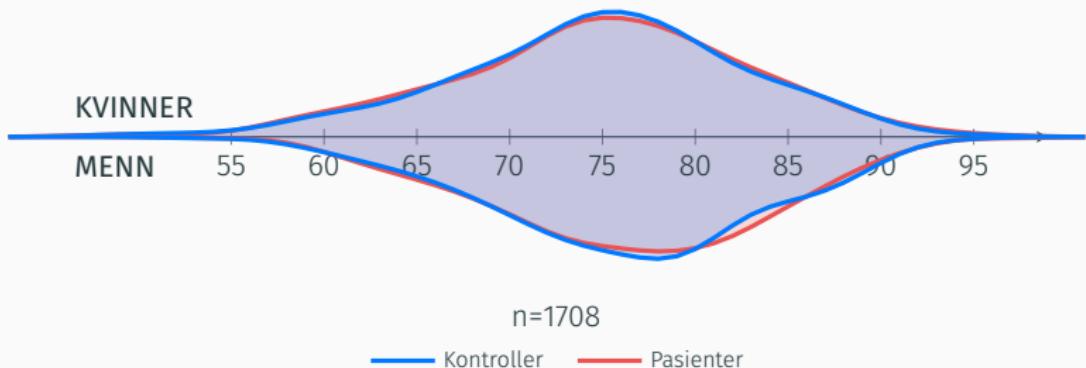


AI i hjernehforskning: Forklarbar AI

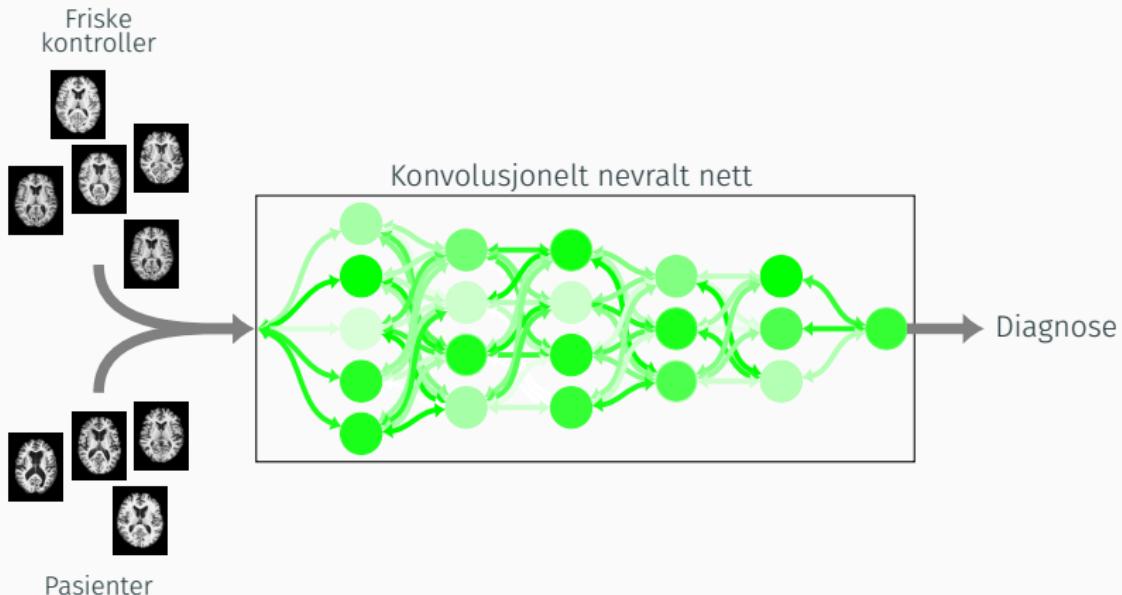




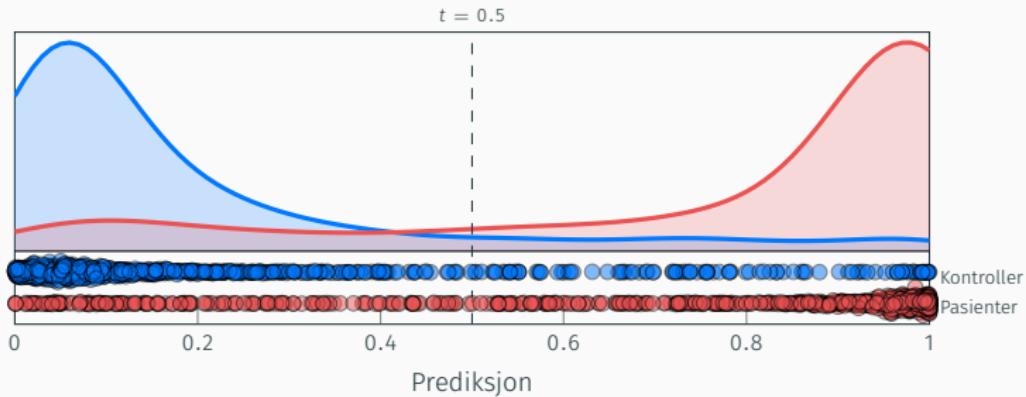
AI i hjernehforskning: Forklarbar AI



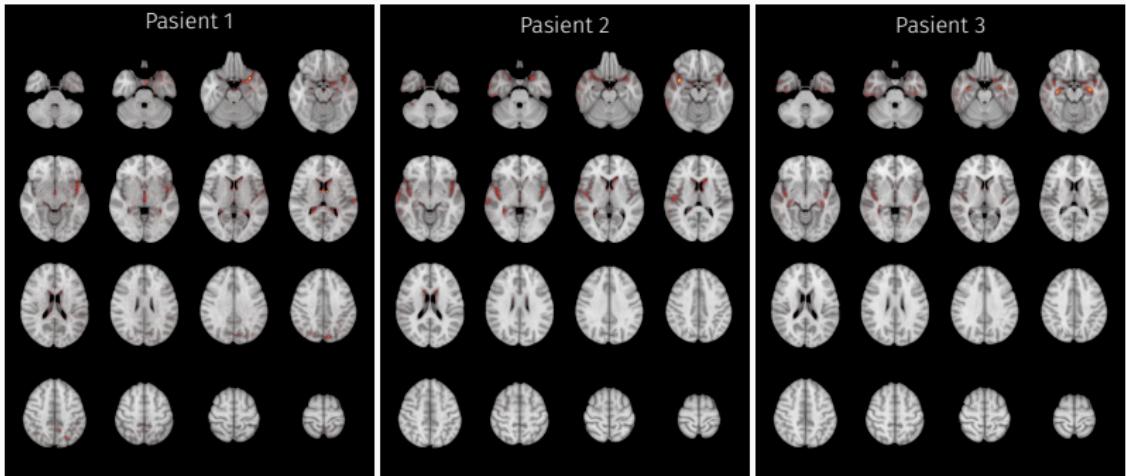
AI i hjernehjerneforskning: Forklarbar AI



AI i hjernehjerneforskning: Forklarbar AI

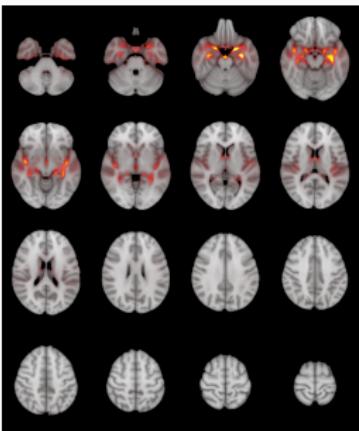


AI i hjernehjerneforskning: Forklarbar AI

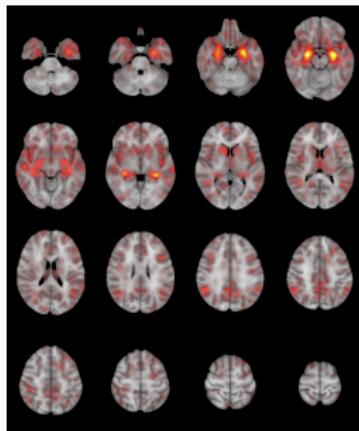


AI i hjernehjerneforskning: Forklarbar AI

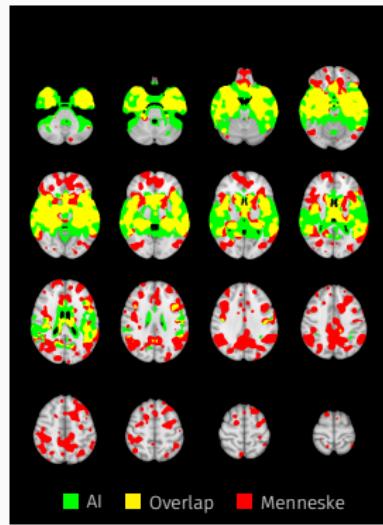
AI



Menneske



AI i hjernehjerneforskning: Forklarbar AI



Forklarbar AI gir oss muligheten til å forstå hvordan modellene opererer.

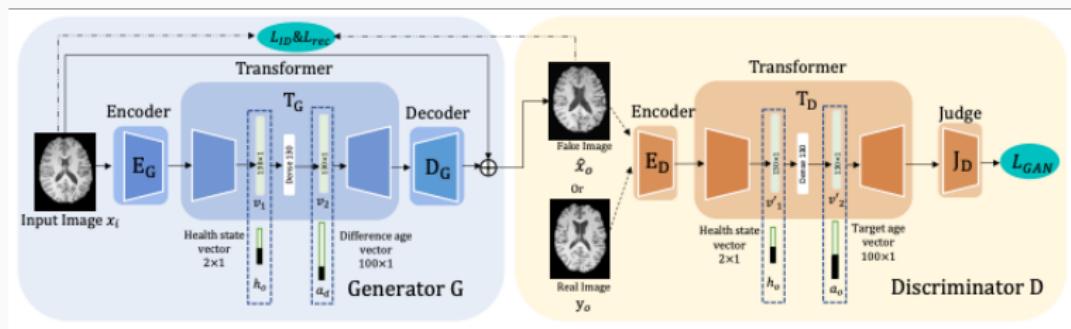
- Kan de lære oss noe?
- Muliggjøre kliniske implementasjoner?

AI i hjernehjerneforskning: Generativ AI

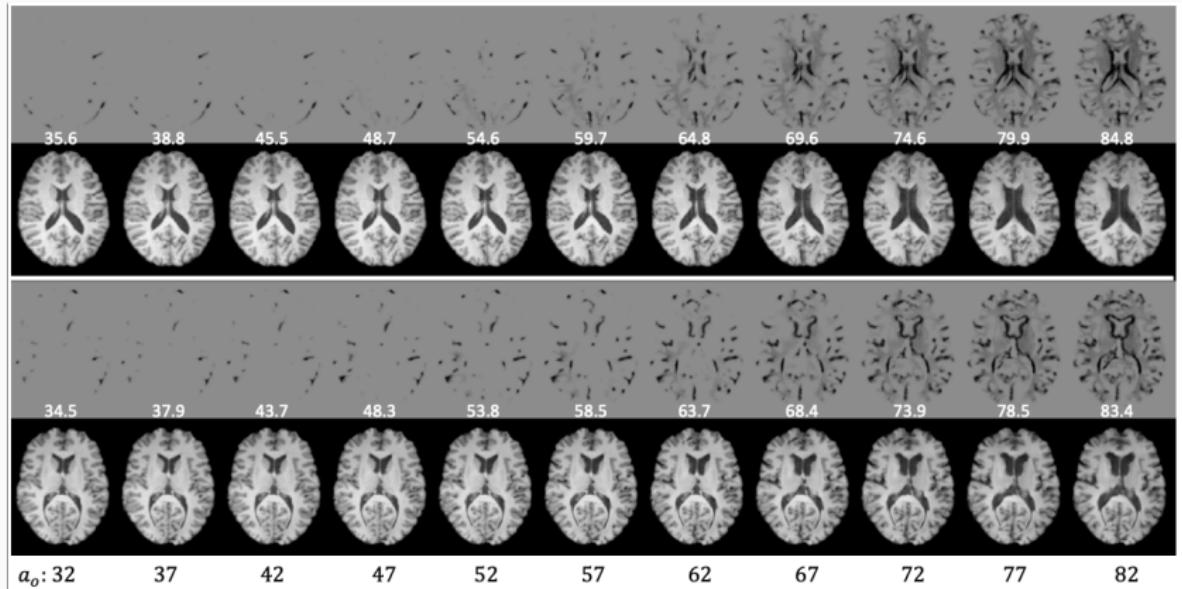


Imagen

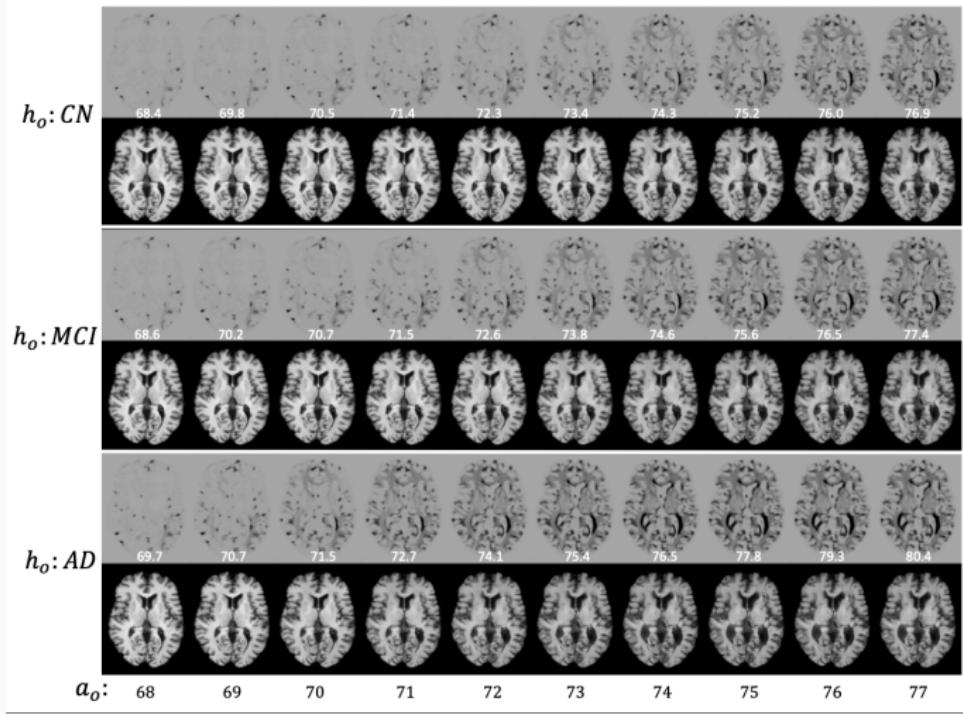
AI i hjernehjerneforskning: Generativ AI



AI i hjernehjerneforskning: Generativ AI



AI i hjernehjerneforskning: Generativ AI



AI i hjernehforskning: Oppsummering

Kan AI bidra til et paradigmeskifte i hjernehforskning?

- Tidligere hjernehforskning (og annen forskning) har i stor grad vært drevet av menneskeskapte hypoteser og hypotesetesting.
- AI kan bidra til å oppdage mønstre som mennesker ikke ser, og dermed generere hypoteser basert på data.
- Dersom man både genererer **og** tester hypoteser i data må man være særs forsiktig.
- Hva er vitsen med kunnskapen dersom mennesker ikke kan forstå den?

Forsker overrasket: Et overveldende flertall ville sjekket hjernehelsen

9 av 10 vil vite om sin risiko for hjernesykdom, viser ny studie fra FHI.

Forskning på og med AI handler ofte om metodeutvikling for å løse generelle problemer. Disse generelle metodene implementeres deretter i spesifikke anvendelser som er utenfor de originale forskernes kontroll.

- Har forskerne et ansvar dersom metodene deres brukes til negative formål, f.eks. autonome våpen eller overvåkningssystemer?

Hvordan ivareta hensynet til enkeltpersoner når vi bruker data til å trenere maskinlæringsmodeller i forskningsøyemed?

- ML er avhengig av store mengder data, ofte uten at man vet hva som trengs på forhånd. Dette står i kontrast med prinsippet om å ikke samle inn mer data enn nødvendig.
- Data behandles som regel med stor forsiktighet, men det er vanlig praksis for forskningsgrupper å dele ferdigtrente modeller. Inneholder disse personopplysninger?

En trenet modell vil anvende mønstre basert på det den ser i treningsdataene. Hva om disse dataene er skjeve?

- ML er ekstremt gode på å finne mønstre, også når disse ikke er de vi ønsker oss, f.eks. basert på variabler som kjønn, alder, eller opphav.
- Man finner ofte at modeller diskriminerer på bakgrunn av disse variablene, også når de ikke finnes eksplisitt i treningsdataen.
- Det er fare for at AI viderefører og forsterker skjevheter i samfunnet.
- På den annen side, gjennom modellene vil skjevhetene bli eksplisitte, kanskje dette gjør de lettere å håndtere?

AI og forskningsetikk: Forklarbarhet

Moderne ML modeller er iboende så kompliserte at det er vanskelig for mennesker å forstå hvordan de fatter slutninger. Kan vi leve med det?

- Er vi egentlig så gode til å forstå hvordan andre mennesker fatter slutninger?
- Forklarbarhet er et aktivt forskningsområde.

AI skiller seg fra mange andre fagfelt ved at mye av den beste forskningen foregår hos kommersielle aktører (Facebook, Google). Det er også disse som sitter på de største datamengdene. Deres hovedfokus er å ivareta sine kommersielle interesser.

- Er dette en selvforsterkende effekt?
- Hvordan kan forskningsgrupper holde seg relevante?
- Hvordan kan samfunnet sikre at forskning på AI kommer alle til gode?

AI og etikk: Verdens ende?

Ledende forskere mener at AI er en eksistensiell trussel mot menneskeheten. Er det grunn til bekymring?

- Hvordan veie oppsidene av AI mot nedsidene?