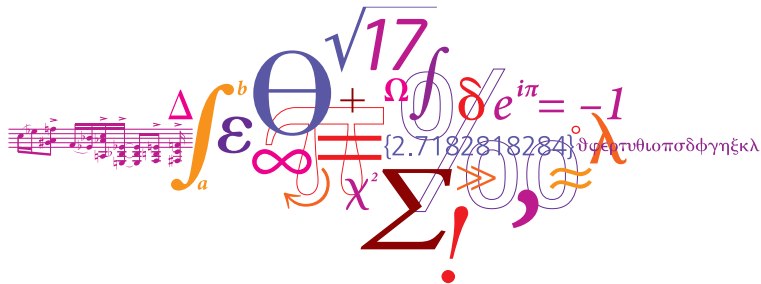


Kunstig intelligens og sociale robotter

Thomas Bolander, DTU Compute, Danmarks Tekniske Universitet

16. december 2019



Lidt om mig selv

Thomas Bolander

- Professor i logik og kunstig intelligens på **DTU Compute**.
- **Aktuel forskning:** Sociale aspekter af kunstig intelligens.
- Medlem af kommissioner m.m. indenfor kunstig intelligens og dets samfundsmæssige aspekter, herunder *SIRI-kommissionen* og *TechDK-kommissionen*.
- H.C. Ørsted sølvmedalje for fremragende videnskabsformidling, 2019.
- Med-arrangør og videnskabelig rådgiver for *Science & Cocktails*.

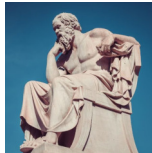


Ny bog (november 2019)



“Gribende page-turner” ★★★★★ Altinget

Min karriere(om)vej



Hvordan undgår vi at
fremtidens robotter
og computere bliver
sociale analfabeter?

Hvad er kunstig intelligens?

John McCarthy, den kunstige intelligens' fader, definerer:

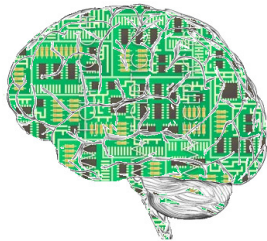
*“Artificial intelligence is the **science and engineering of making intelligent machines, especially intelligent computer programs.**”*

(McCarthy, 1956)

Problem: Utroligt mange forskellige former for intelligens, og på meget forskellige niveauer.

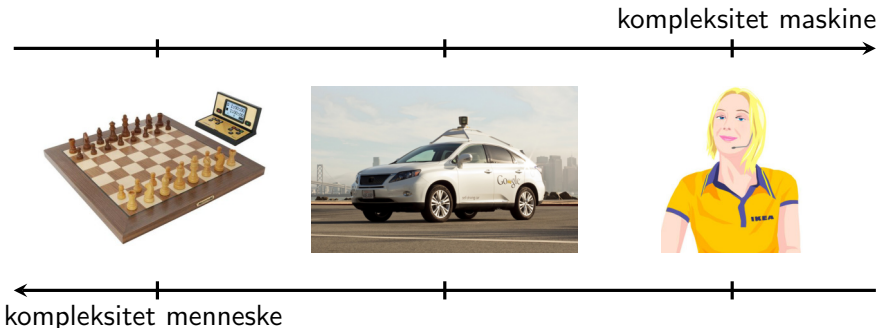


John McCarthy, 2006



Karakteristika ved kunstig intelligens-systemer i dag

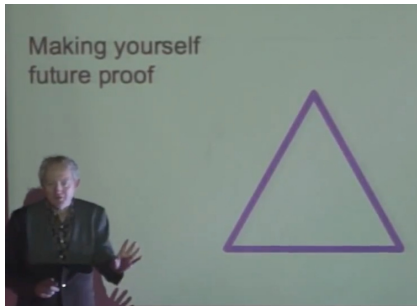
- **Specialiserede systemer:** Løser veldefinerede, velafgrænsede problemer.
- **Tommelfingerregel:** Jo mere veldefineret og velafgrænset et problem er, des lettere er det at lave kunstig intelligens som kan løse det.
- **Revolutionen skyldes i høj grad beregningskraft og data:** mere end udvikling af nye metoder med højere kognitive evner.



De 3 sværeste problemer i kunstig intelligens



Carl Frey, 20. april 2017
*Udd.- og Forskningsministeriets
Årsmøde 2017, Kolding*



Toby Walsh, 18. marts 2017
Science & Cocktails, København

Begge har **social intelligens** blandt de 3 menneskelige evner som er sværest at simulere på computere og i robotter.

Social intelligens: Evnen til at forstå andre og den sociale kontekst, og dermed at interagere succesfuldt med andre.

To hovedparadigmer i kunstig intelligens

Det symbolske paradigme (1950–): Simulerer menneskelig symbolsk, sproglig, bevidst ræsonnering. Søgning, planlægning, logisk ræsonnering. **Eks:** skakcomputer.



robust, forudsigelig, forklarlig



nøje afgrænsede evner



fleksibel, læring



aldrig 100% forudsigelig/fejlfri

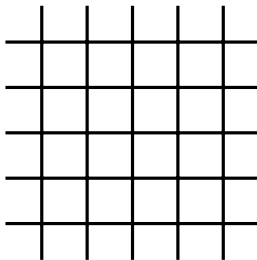
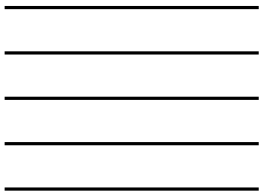
Statistisk læring (1960–): Simulerer typisk de fundamentale fysiske processer i den menneskelige hjerne. Kunstige neurale netværk m.m. **Eks:** billedgenkendelse.

symbolsk



statistisk læring

Billedgenkendelse med statistisk læring: læring og genkendelse af geometriske mønstre



Billedgenkendelse med statistisk læring: læring og genkendelse af geometriske mønstre

Med din sidemakker: Prøv at beskrive forskellen på katte og hunde i rent geometriske termer, altså via geometriske former/kendetegn, størrelsesforhold og relative positioner. Alt skal i sidste ende reduceres til geometri, så du kan ikke sige "øjnene er X " uden at du så også bliver nødt til at beskrive i geometriske termer hvad et øje er (hvordan man geometrisk kan genkende det).

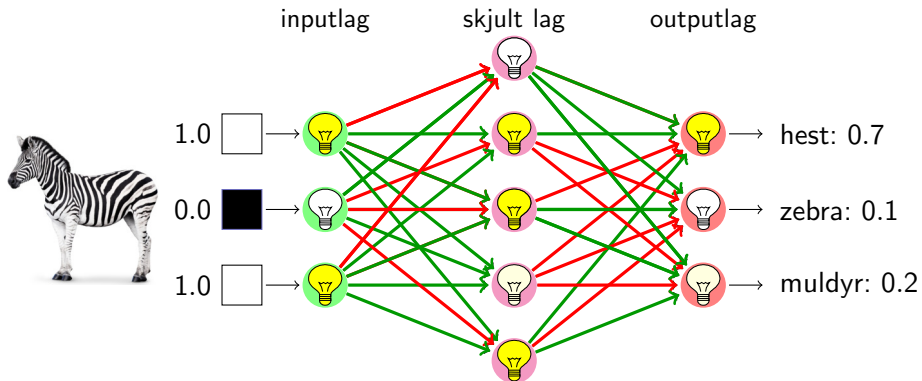




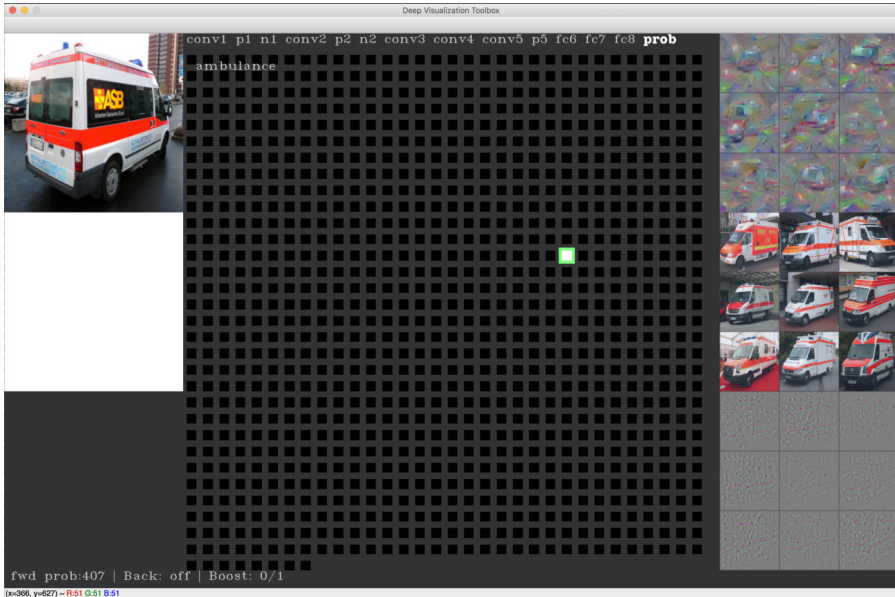
Viktigste teknik i statistisk læring: Kunstige neurale netværk

Et (kunstigt) neuralt netværk producerer et **output** (et eller flere tal) fra et **input** (et eller flere tal). Output er ofte en **klassifikation**: hvilken klasse tilhører inputtet, fx er det en kat eller en hund?

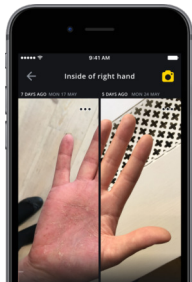
Neuron, fra Wikipedia: "Synaptic signals may be excitatory or inhibitory, increasing or reducing the net voltage that reaches the soma."



Demo af kunstige neurale netværk

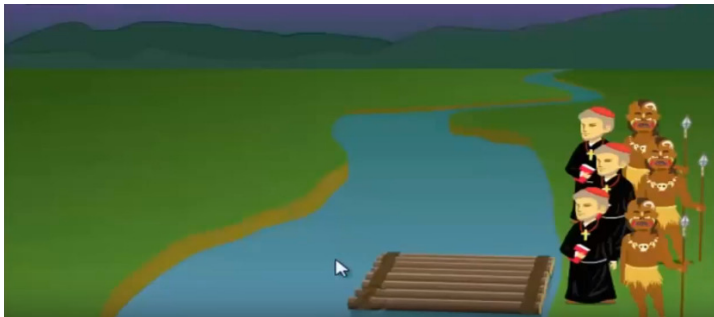


Anvendelseksemppler af kunstige neurale netværk i Danmark

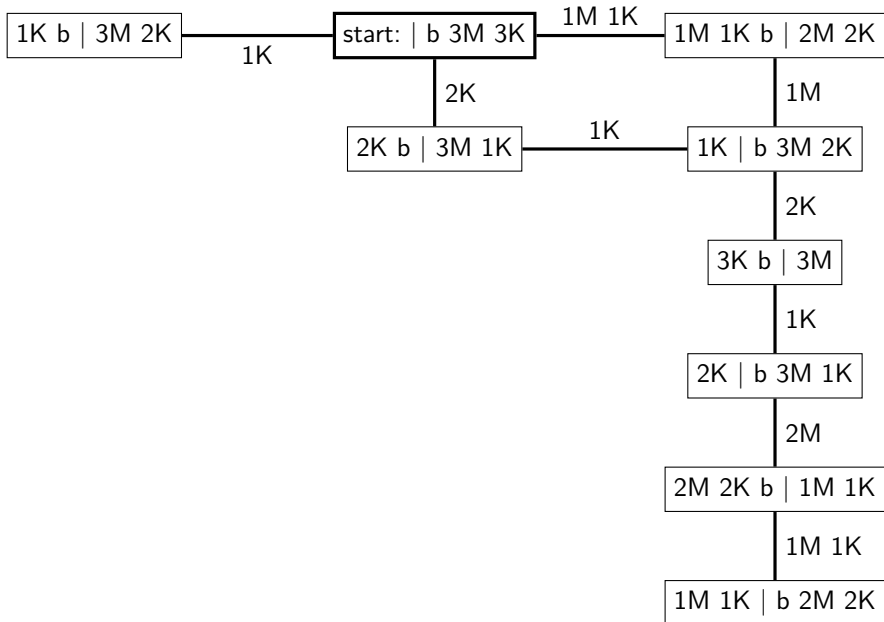


Søgning i symbolsk AI: Missionærer og kannibaler

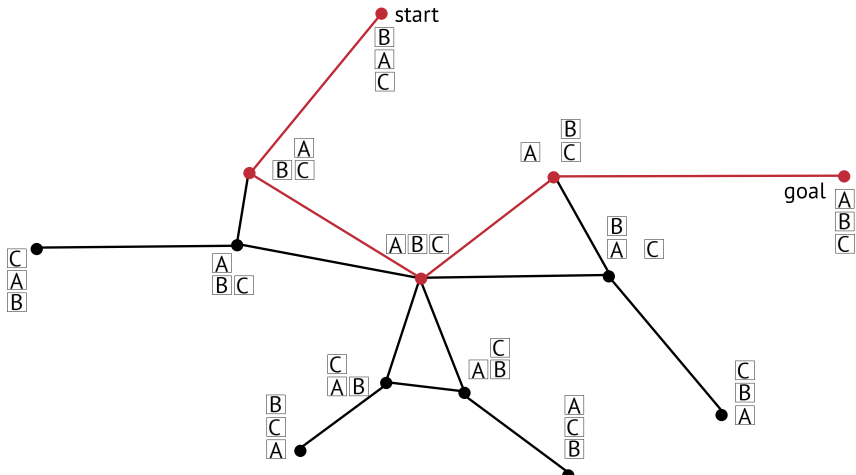
Med din sidemakker: Tre missionærer og tre kannibaler er på en flodbred. De har en tømmerflåde med plads til to personer. Find en måde at få alle fra den ene til den anden flodbred uden at antallet af kannibaler på noget tidspunkt overstiger antallet af missionærer, hverken på den ene eller den anden bred. (Når båden sejler fra den ene til den anden bred, skal alle stige ud inden båden kan fyldes op igen. Båden kan ikke sejle uden der er nogen i den.)



Missionærer og kannibaler



Automatiseret planlægning/søgning



Anvendelseksempel af automatiseret planlægning: Amazon-robotter

http://www2.compute.dtu.dk/~tobo/amazon_kiva.mp4

Social kunstig intelligens: Hvorfor?

- Fleksibel og naturlig interaktion med mennesker.
- Forklarlighed: AI-systemer der kan gøre sig forståelige for mennesker (skabe tillid).

Eksempel på nødvendigheden af social intelligens. Kritiske brugere af hospitalsrobotter i USA.

- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*

(Colin Barras, New Scientist, vol. 2738, 2009)



TUG hospitalsrobot

Social intelligens: Hvad er det?

Et kognitionspsykologisk eksperiment med en dreng på 18 måneder. Han har ikke fået nogen instruktioner. (Warneken & Tomasello, 2006)

http://www2.compute.dtu.dk/~tobo/children_cabinet.mpg

Kunstig intelligens som kan løse samme problem som den lille dreng?



dreng:
gå til skab



dreng:
åben skab



voksen:
gå til skab



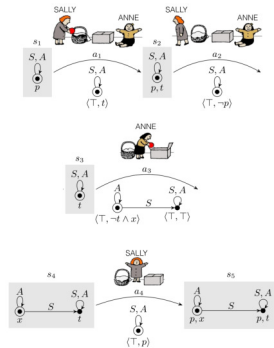
voksen:
læg papirer



Min forskning i robotter med social intelligens



statistisk læring



symbolsk

Løser kognitive opgaver: **false-belief tasks** af vilkårlig orden. Mennesker løser lette af disse når man er 4, men svære først når man er 20.

- **Statistisk læring** (perception): ansigtsgenkendelse, objektgenkendelse, skelet-tracking, tale-til-tekst.
- **Symbolisk AI** (højere kognition): planlægning, intentioner, logisk ræsonnering, perspektiv-tagning.

APPENDIKS: MERE OM NEURALE NETVÆRK

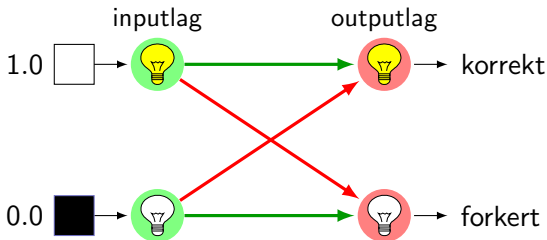
Hvad er et neuralt netværk?

Et neuralt netværk producerer et **output** (et eller flere tal) fra et **input** (et eller flere tal). Ofte brugt til **klassifikation**. Hvilken klasse tilhører inputtet? Er det fx en kat eller en hund (input er billede).

Eksempel. Genkend mønsteret blandt alle 1x2 pixel billeder i sort-hvid (inklusive gråtoner).



blandt alle 1x2 pixel billeder i



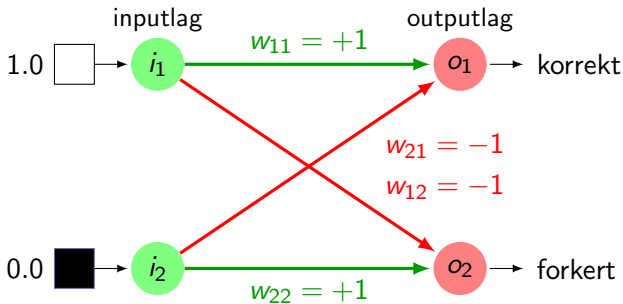
Neuron, fra Wikipedia: "Synaptic signals may be excitatory or inhibitory, increasing or reducing the net voltage that reaches the soma."

Neurale netværk med talvægte

Vi ønsker stadig at genkende mønsteret blandt alle 1x2 pixel billeder i sort-hvid.



blandt alle 1x2 pixel billeder i sort-hvid.

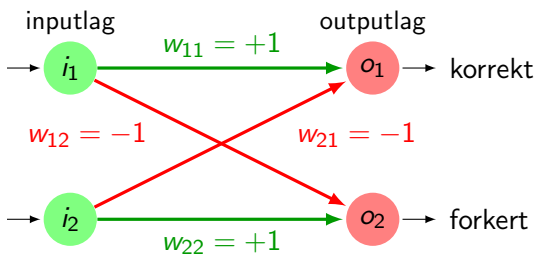



$$\begin{aligned}o_1 &= w_{11}i_1 + w_{21}i_2 \\ &= i_1 - i_2\end{aligned}$$

$$\begin{aligned}o_2 &= w_{12}i_1 + w_{22}i_2 \\ &= -i_1 + i_2\end{aligned}$$

Med input $(i_1, i_2) = (1, 0)$ fås output $(o_1, o_2) = (1, -1)$, så inputtet bliver angivet som "korrekt". Med input $(i_1, i_2) = (0, 1)$ fås output $(o_1, o_2) = (-1, 1)$, så inputtet bliver angivet som "forkert".

Håndtering af støj



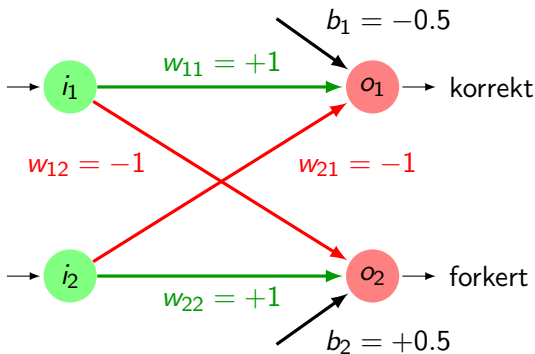
Støjfyldt, men stadig "korrekt", input: 0.8 
 0.2

Med dette input (altså $(i_1, i_2) = (0.8, 0.2)$) fås output:

$$\begin{aligned}(o_1, o_2) &= (w_{11}i_1 + w_{21}i_1, w_{12}i_1 + w_{22}i_2) \\ &= (i_1 - i_2, -i_1 + i_2) \\ &= (0.6, -0.6)\end{aligned}$$

Første udfordring: Hvad med dette input? 0.5 
 0.5

Løsning til første udfordring: Introduktion af bias



Nye formler:

$$\begin{aligned}o_1 &= w_{11}i_1 + w_{21}i_2 + b_1 \\ &= i_1 - i_2 - 0.5\end{aligned}$$

$$\begin{aligned}o_2 &= w_{12}i_1 + w_{22}i_2 + b_2 \\ &= -i_1 + i_2 + 0.5\end{aligned}$$

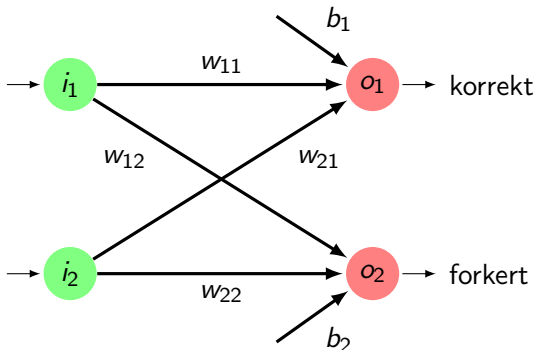
Med input $\begin{matrix} o_1 = 0.5 \\ o_2 = 0.5 \end{matrix}$ fås nu

$$(o_1, o_2) = (i_1 - i_2 - 0.5, -i_1 + i_2 + 0.5) = (-0.5, 0.5).$$



Genkendelse af flere forskellige mønstre

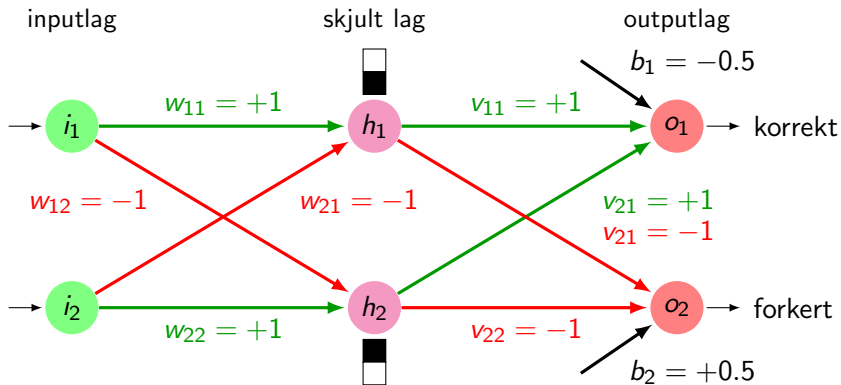
Antag vi betragter både $\begin{matrix} 1 & \square \\ 0 & \blacksquare \end{matrix}$ og $\begin{matrix} 0 & \blacksquare \\ 1 & \square \end{matrix}$ som korrekte mønstre.

Anden udfordring: Hvordan justerer vi vægtene?



Løsning til anden udfordring: Skjulte lag

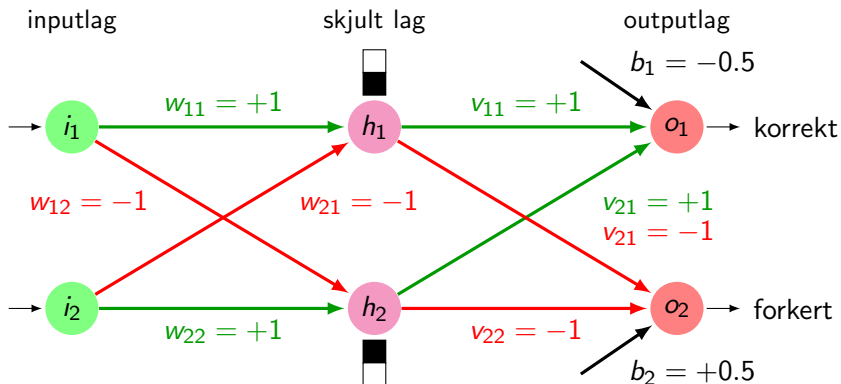
Vi har skabt en output-neuron til at genkende mønsteret  og en anden til at genkende . Så hvad med at kombine outputtet fra to sådanne neuroner ind i en tredje neuron? Dette giver os et neuralt netværk med et tredje lag, kaldet et **skjult lag**.



Tredje udfordring: Hvad sker der nu med input  ?

Løsning til tredje udfordring: Aktiveringsfunktion

Aktiveringsfunktion σ : Rund negative tal op til 0 (kaldes **ReLU**-aktiveringsfunktionen).



Lad $(i_1, i_2) = (1, 0)$. Så gælder

$(h_1, h_2) = (\sigma(i_1 - i_2), \sigma(-i_1 + i_2)) = (\sigma(1), \sigma(-1)) = (1, 0)$. So

$(o_1, o_2) = (\sigma(h_1 + h_2 + b_1), \sigma(-h_1 - h_2 + b_2)) = (\sigma(0.5), \sigma(-0.5)) =$

$(0.5, 0)$. Outputtet er "korrekt". Symmetrisk med $(i_1, i_2) = (0, 1)$.

Øvelse.

- a. Betragt en vandret line i øverste række af et 2x2 pixel billede:



Konstruér et neuralt netværk der korrekt kan identificere dette mønster ved at bruge ideerne fra de foregående slides. Afprøv dit neurale netværk på et par forskellige inputs (både korrekte og forkerte), for at tjekke at det virker.

- b. Modificér nu dit neurale netværk så det korrekt kan identificere vandrette linjer uafhængigt af deres position, altså som identificerer begge følgende mønstre som korrekte:



Afprøv dit neurale netværk på forskellige inputs.

Træning af et neuralt netværk

Antallet af lag, antallet af neuroner i hvert lag og aktiveringsfunktionen bliver alle besluttet af menneskelige programmører. Disse kaldes **hyperparametre**. Neuronvægtene kaldes derimod **parametre**, og de er disse der bliver automatisk justeret gennem træningen af det neurale netværk.

Neuronvægtene har til at starte med tilfældige værdier, og bliver så justeret igennem **træning** af det neurale netværk:

1. Send et eksempel-input ind i det neurale netværk.
2. Observér fejlen på outputtet (forskellen mellem det faktiske og det forventede output).
3. Justér vægtene baglæns gennem netværket for at reducere fejlen på det specifikke input (**backpropagation**).
4. Gentag indtil man får et tilfredsstillende resultat på test-sættet (forskelligt fra trænings-sættet).