

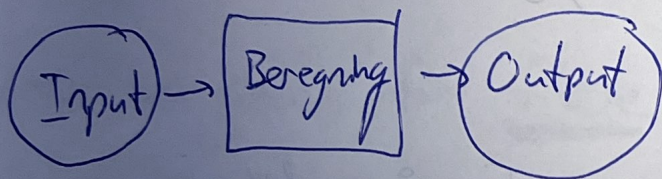
# Forelesning 6:

## LSTM - Long Short-term Memory, men først...

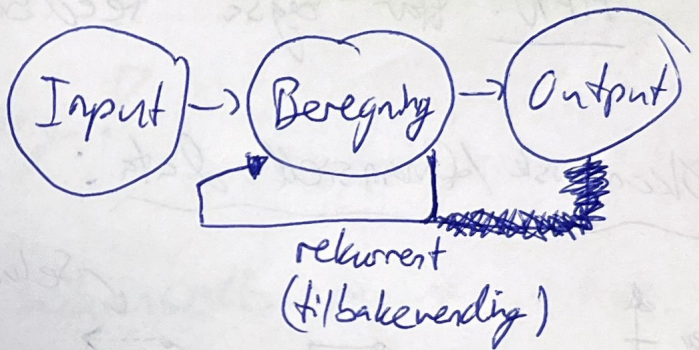
### • Recurrent neural net works

- RNN
- Recurrent = tilbakevendende

### MLP (Multi layer perceptron)



### RNN



### Hvordan fungerer RNN?

- MLP: Informasjonen går én vei: input → output
- RNN: Har tilbakevendinger.

Obs: Å forstå RNN er en slags "stepping stone" for å forstå LSTM (og Transformers)



## Til nå:

- Sett på neurale nettverk som har et gitt form på input.

↳ 28x28 piksel bilder (MNIST)

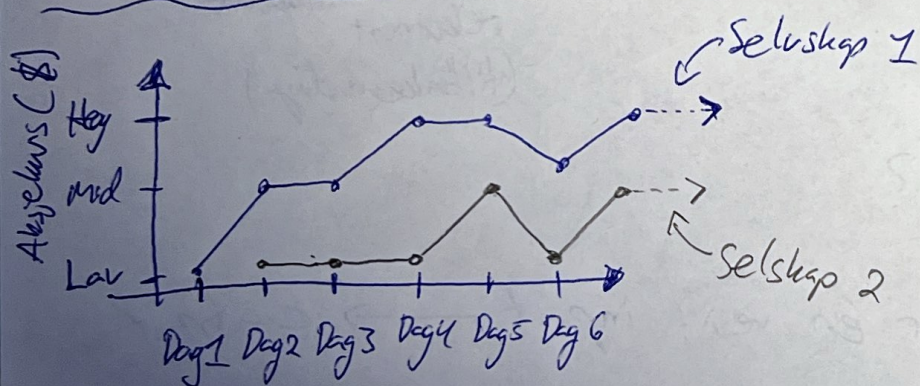
↳ California Housing, Fraud, ...

## RNN's:

- Har akt.funksjoner, bias, vekter, skjulte/input/output lag.

MEN: Har også "feedback loops"

## Økonomisk / Finansiell data:

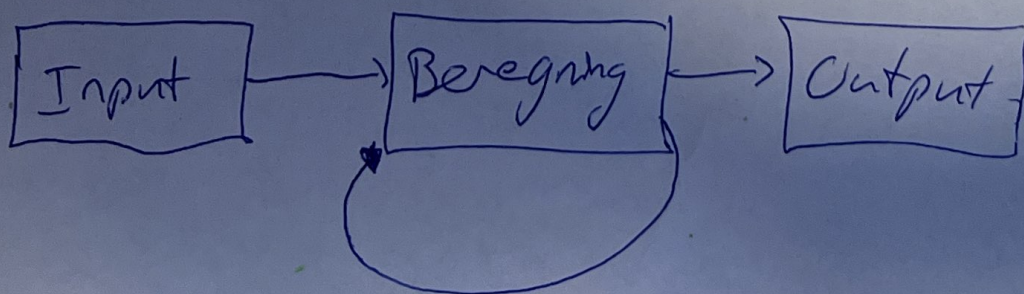
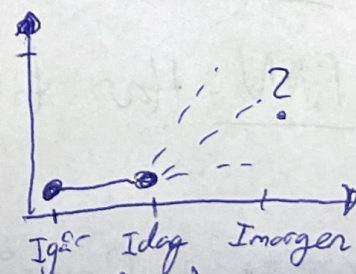


For å predikere fremtiden her, ønsker vi å se på dataene fra foregående dager, for å si noe om morgendagen.

## RNN - et enklere eksempel

• Vi ser på Igår, idag og i morgen.

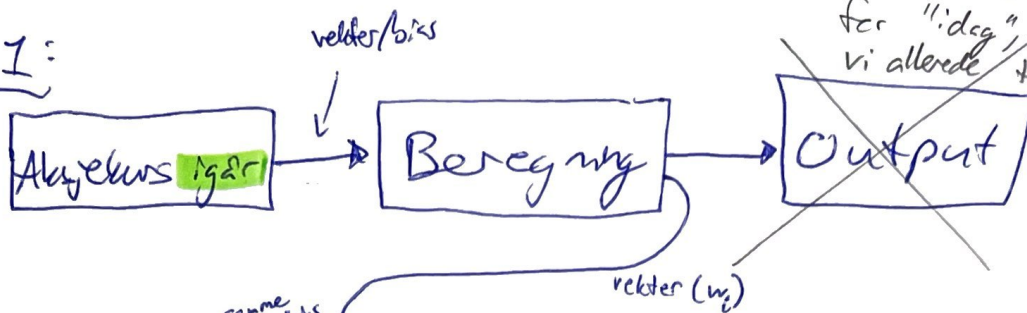
→ To input, av sekvensell form (igår og idag)





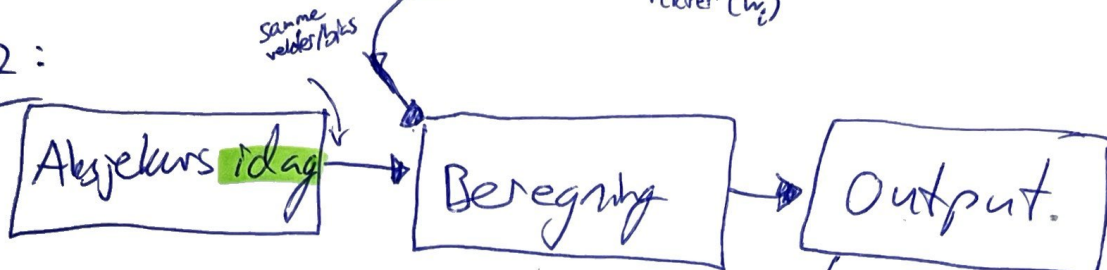
Oh, hva sier egentlig

Steg 1:



Vi ignorerer dette outputet, fordi det er for "i dag" men det kjennes vi allerede til!

Steg 2:



Abjekursen i morgen:

Dette kan gjøres for flere steg!

Viktig:

Uansett hvor mange sekvensielle input vi har, er det samme vektor/biases til hvert input. Derfor blir ikke antall vektor/biases som følge av dette. [Derfor kan det "takle" sekvenser av ulike lengde]

Men, jo flere steg vi tar. Jo hardere/vanskeligere er det å trene nettverket.

Det fører (ofte) til vanishing gradient problem.  
(exploding — " — ).

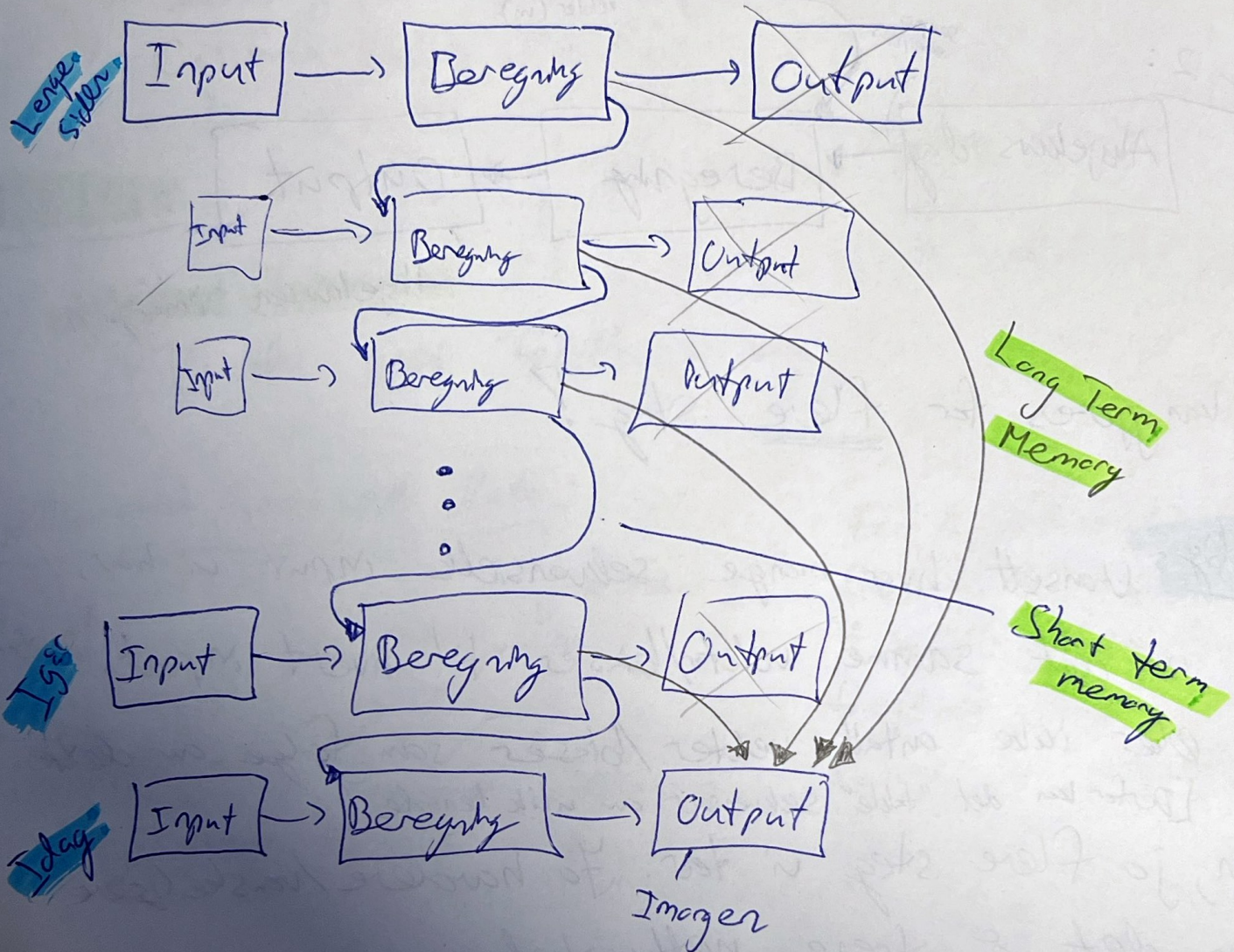
↳ At gradienten i gradient descent, som skal finne minimum av loss funksjonen, enten blir veldig stor (exploding) eller veldig liten (vanishing), som følge av vektorene  $w_i$  fra steg 1  $\rightarrow$  steg 2  $\rightarrow$  ...



Løsning på varishing/exploding gradient problem:

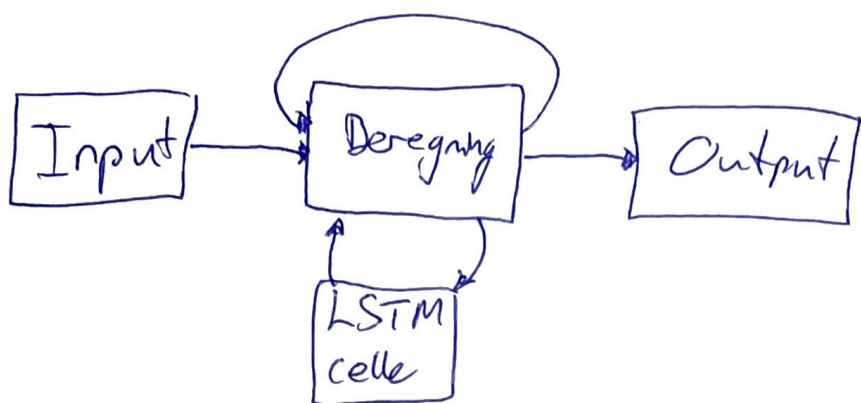
# LSTM

- En RNN san loser vanishing/~~gradient~~<sup>exploding</sup> gradient problemet.



- LSTM bruker to ulike veier til output, én for short-term (slik som vanlig RNN) og én for Long-term.





Så hva er en LSTM-celle?

- Hjelper oss i kunne bruke langer sekvenser enn vanlige RNN.
- Består av tre deler:

Forget	Input	Output
gate	gate	gate

Forget gate: En verdi mellom 0 og 1 (via sigmoid), som bestemmer hvor mye fra tidligere steg skal bevares / kan glemmes.

1 = "husk alt", 0 = "glem alt"

Input gate: Opererer på samme signal som "forget gate", men her er målet å bestemme hvilke ny info skal inn i vår LSTM. Output av input gate (mellom 0 og 1) ganges med output av tanh, som produserer nye verdier som skal legges til i tidligere state.

Output gate: Bestemmer hvor mye/hva som skal sendes videre fra LSTM celler og til neste lag i nettverket.