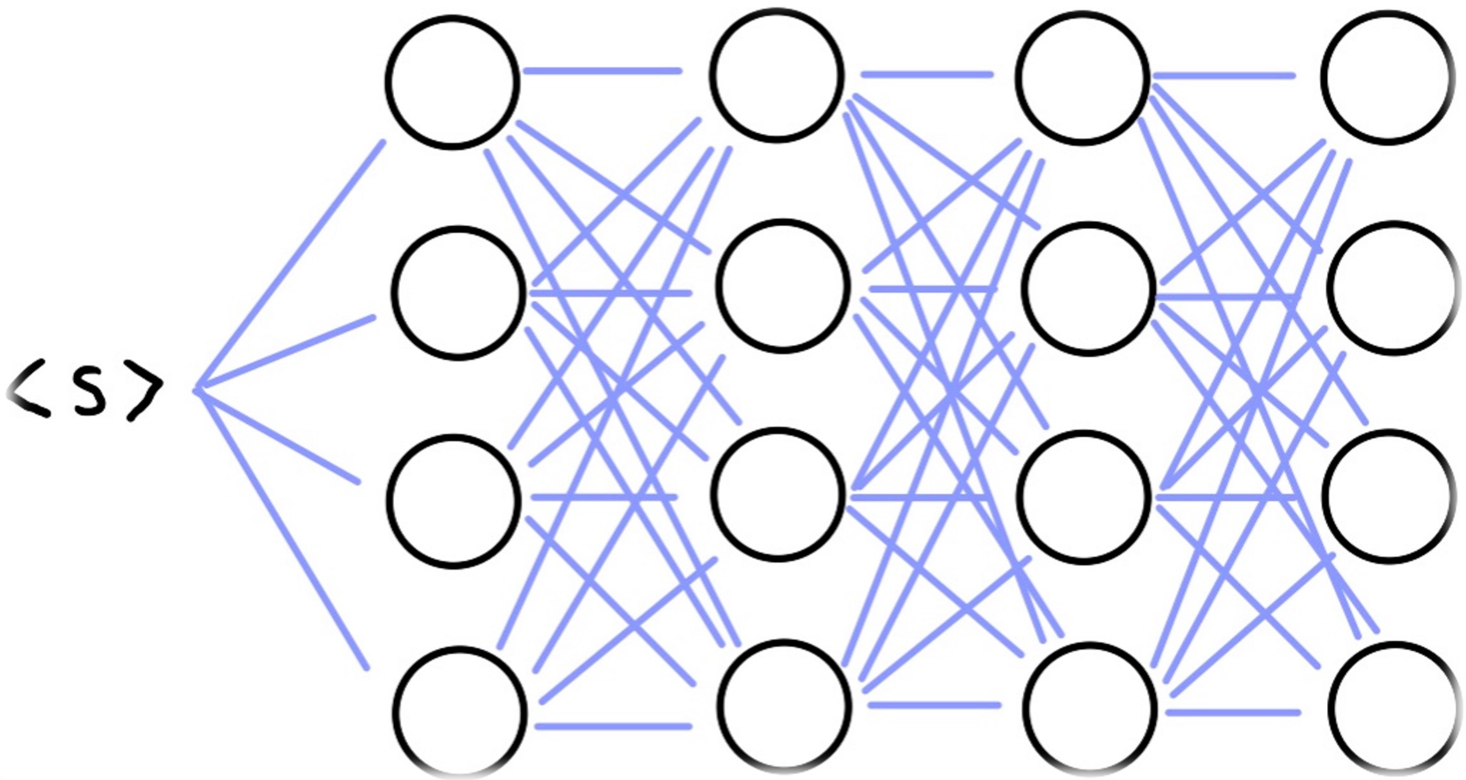
# VITERBI



# VITERBI-DEKODING

- Dekodingsalgoritme: Beregner den mest sannsynlige sekvensen av skjulte tilstander ved hjelp av en HMM
- Lineær kompleksitet: Behandler hver sekvens ord for ord
  - *Mye* raskere enn eksempelet fra forrige gang!
- Beregner sannsynligheten for neste tilstand $t$ basert på den forrige tilstanden $t-1$

# VITERBIFORMELEN

emisjonsmodellen

$$v_t(x) = \max_{1 \le y \le N} v_{t-1}(y) \, P(s_t = x | s_{t-1} = y) \, P(o_t | s_t = x)$$

Velg den mest sannsynlige ordklassen *y* for forrige ord

transisjonsmodellen

# KORT SAGT:

- Første steg:
  - Beregn sannsynligheten for starttilstandene ved transisjon * emisjon
  - Ta vare på den høyeste sannsynligheten ved hjelp av en *backpointer*
- *Resten av stegene:*
  - Beregn sannsynligheten for tilstandene ved transisjon * emisjon * forrige resultat
- Siste steg:
  - Følg den mest sannsynlige stien av skjulte sekvenser *baklengs*, via backpointers

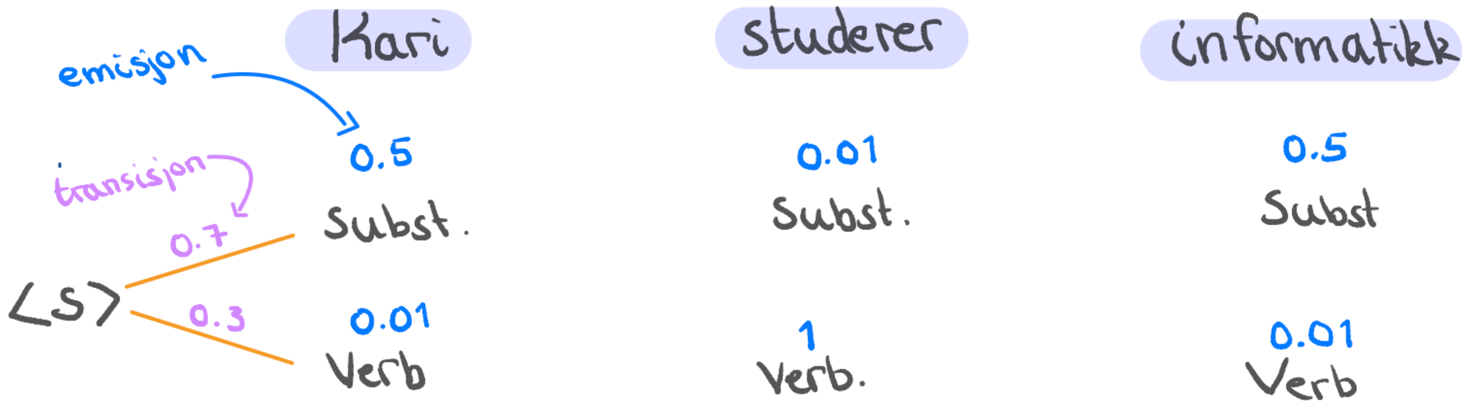
    \- Dette er svaret vårt!👏

# BEREGNINGSEKSEMPEL

Hva er den mest sannsynlige sekvensen av skjulte tilstander for setningen
*Kari studerer informatikk*?

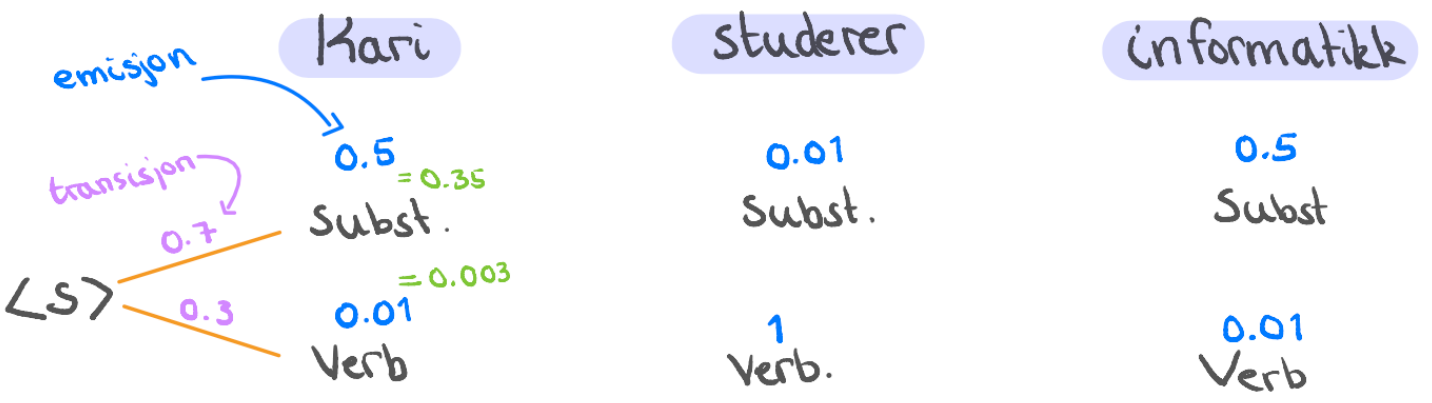
## Steg 1

Finn alle mulige starttilstander <s> og deres sannsynlighetsverdier



## Steg 2

- Finn sannsynligheten for hver mulige starttilstand
- For hver neste tilstand, beholder vi en "backpointer" til den mest sannsynlige forrige tilstand
- Vi lagrer sannsynligheten til den tilstanden som mest sannsynlig tilhører starttilstanden



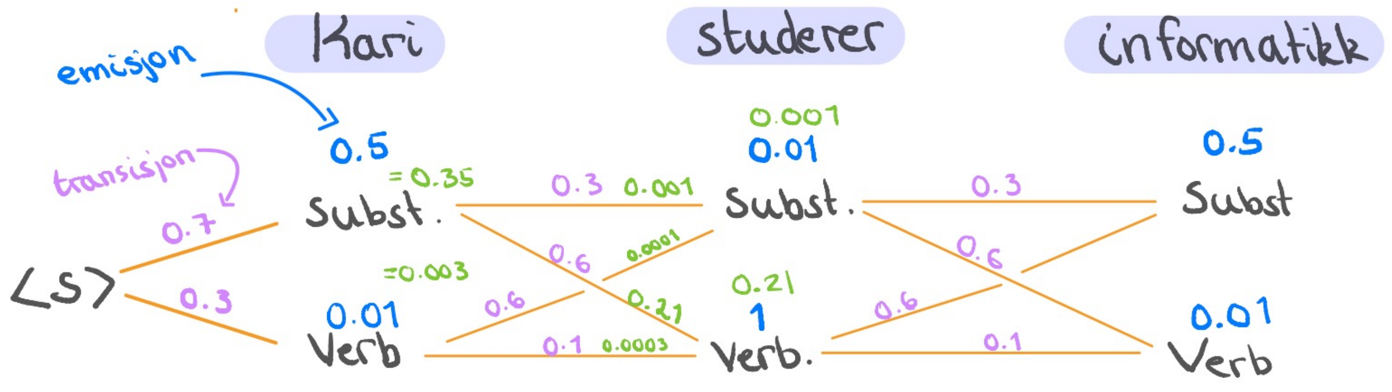

For å finne den mest sannsynlige etterfølgende tilstanden ganger vi transisjon med emisjon.

$P(\text{subst} \mid \text{<start>}) = 0.7 \times 0.5 = 0.35$

$P(\text{verb} \mid \text{<start>}) = 0.3 \times 0.01 = 0.003$

# Steg 3

Gjenta steg 2 for resten av setningen (tilstandene/merkelappene), slik at vi sitter igjen med "stier" av backpointers gjennom setningen

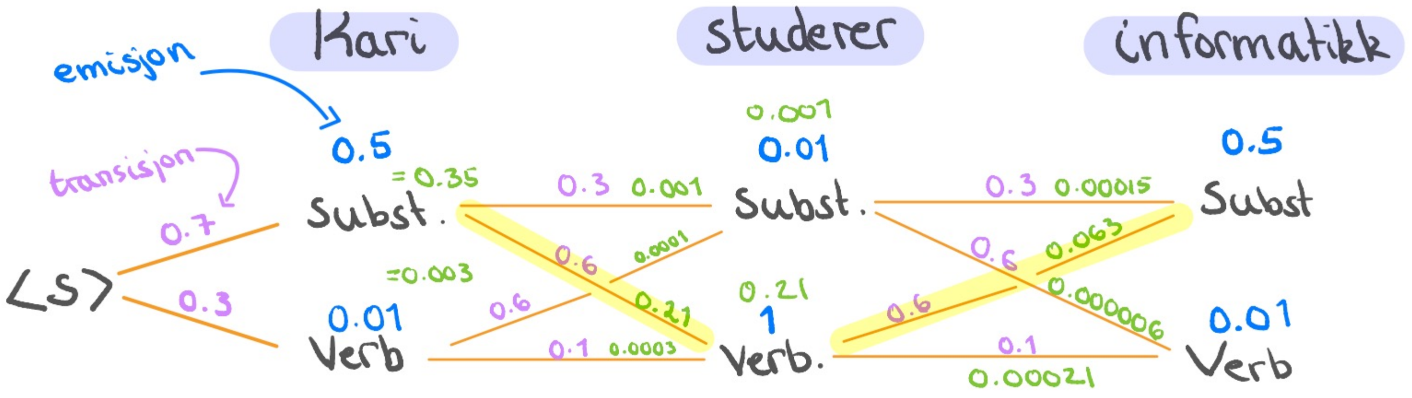



Vi ganger det høyeste resultatet for hver tilstand med utregningen av neste tilstand, som igjen er transisjon x emisjon:

| | |
|---|---|
| $P(subst \mid subst) = 0.35 \times 0.3 \times 0.01 = 0.01$ | $P(subst \mid subst) = 0.001 \times 0.3 \times 0.5 = 0.00015$ |
| $P(verb \mid subst.) = 0.35 \times 0.6 \times 1 = 0.21$ | $P(verb \mid subst.) = 0.001 \times 0.6 \times 0.1 = 0.000006$ |
| $P(verb \mid verb) = 0.003 \times 0.1 \times 1 = 0.0003$ | $P(verb \mid verb) = 0.21 \times 0.1 \times 0.1 = 0.00021$ |
| $P(subst \mid verb) = 0.003 \times 0.6 \times 0.01 = 0.0001$ | $P(subst \mid verb) = 0.21 \times 0.6 \times 0.5 = 0.063$ |

# Steg 4

Til slutt, starter vi bakerst i sekvensen vår, og følger den mest sannsynlige stien til vi er kommet til start. Dette gjør vi ved å velge den stien med det høyeste resultatet for hvert steg.




Den mest sannsynlige sekvensen er: (subst. | Kari) (verb | studerer) (subst. | informatikk)

Utregning i pdf: her