Lecture reviews — Week 06

J.-C. Chappelier & M. Rajman

Laboratoire d'Intelligence Artificielle
Faculté I&C

Week(s 5 &) 6 keypoints

1- what MM and HMM are in details

$\rightarrow$ PoS tags

2- 3 problems

1. $P(w_1^n | \theta) = \sum_{t_1^n \in \mathcal{T}^n} P(w_1^n, t_1^n | \theta)$

2. $\operatorname{Argmax}_{t_1^n} P(t_1^n | w_1^n, \theta)$

3. unsupervized learning : $\operatorname{Argmax}_{\theta} P(\theta | w_1^n)$

Week(s 5 &) 6 keypoints

Week 5:

- what "lemmatization" is
- what "part-of-speech tagging" is
- two hypothesis to transform PoS tagging into "the second problem" of HMMs
- order of magnitude of performances

Week 6:

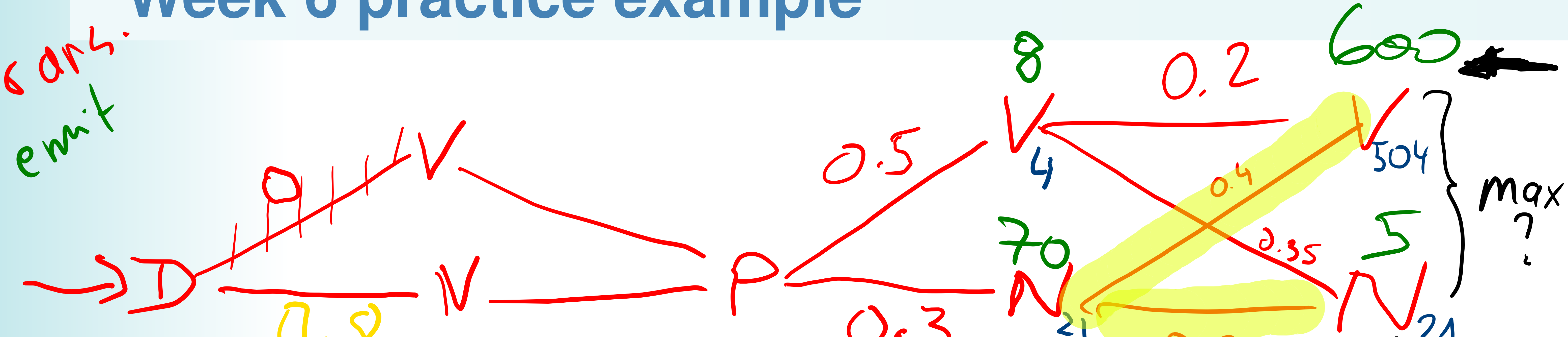
- what an HMM is
- the 3 problems and how it relates to PoS tagging
- Viterbi algorithm : do tagging
- properties of Baum-Welch algorithm

↳ unsupervised learning

Week 6 practice example

• trans.
• emit

4 domain



the cat you saw running

④ What is the most probable tagging (using data provided below)?

```
cat: N (1e-4), V (2e-6)          saw: N (7e-4), V (8e-5)
run: N (3e-6), V (4e-4)          the: D
running: N (5e-6), V (6e-4)      you: P
```

70e-5

600e-6

```
Pi(D) = 0.35    Pi(N) = 0.25    Pi(V) = 0.15    Pi(P) = 0.1
```

P(D|D) = 0	P(N|D) = 0.8	P(V|D) = 0	P(P|D) = 0
P(D|N) = 0.1	P(N|N) = 0.2	P(V|N) = 0.4	P(P|N) = 0.3
P(D|V) = 0.15	P(N|V) = 0.35	P(V|V) = 0.2	P(P|V) = 0.25
P(D|P) = 0.1	P(N|P) = 0.3	P(V|P) = 0.5	P(P|P) = 0

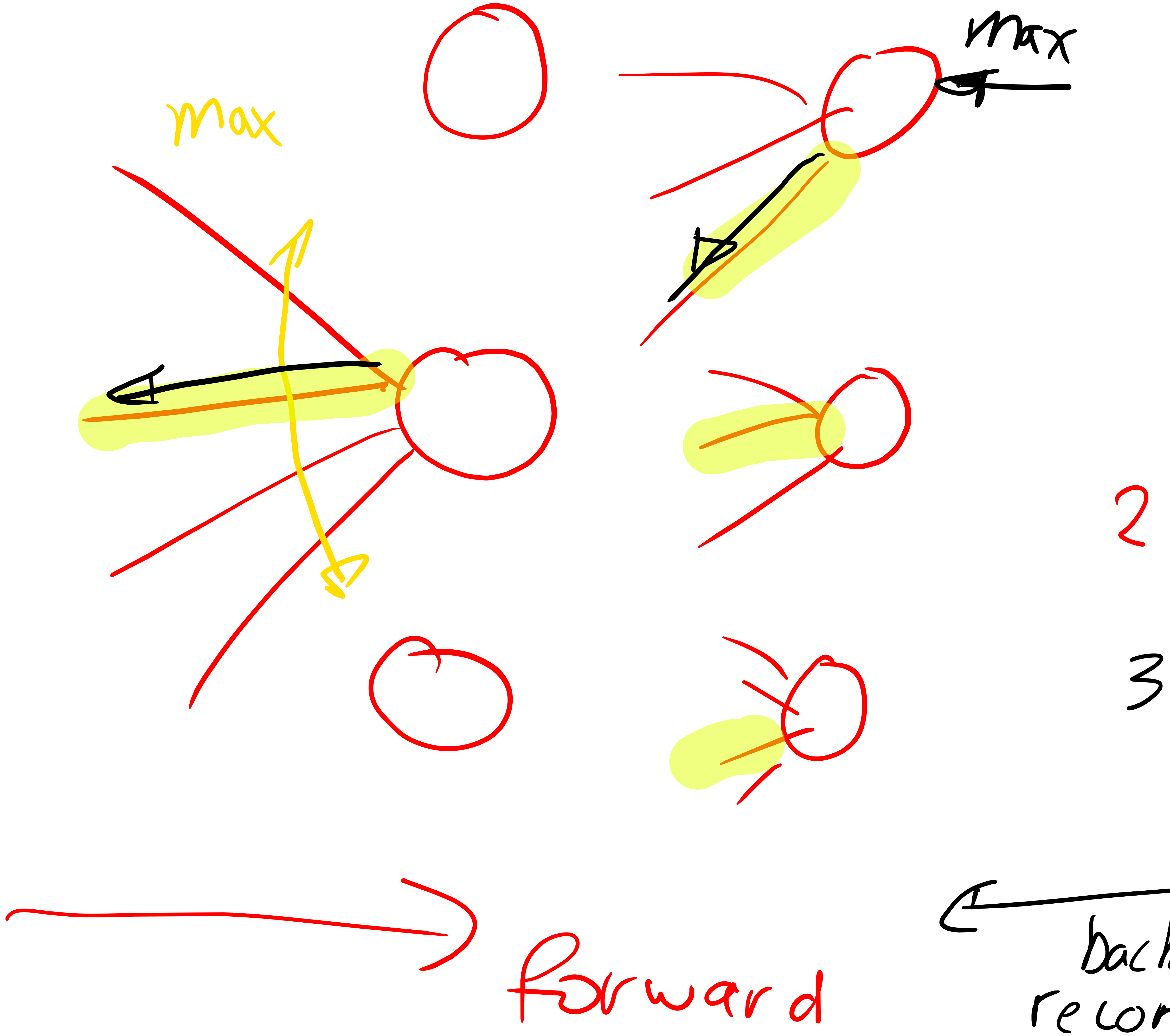
Answ: D N P N V

$$P(D\,N\,P\,N\,N \mid \text{the cat you saw running})$$

$$= \underbrace{P_{init}(D)}_{\text{init}} \cdot \underbrace{P(\text{the} \mid D)}_{\text{emit}} \cdot \underbrace{P(N \mid D)}_{\text{transit}} \cdot P(\text{cat} \mid N) \cdot$$

$$P(P \mid N) \cdot P(\text{you} \mid P) \cdot P(N \mid P) \cdot P(\text{saw} \mid N)$$

$$\cdot\, P(N \mid N) \cdot P(\text{running} \mid N)$$



1 - drawing lattice (incl. init, transit, emit } probs)

2 - forward

3 - backward

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE – LAUSANNE
POLITECNICO FEDERALE – LOSANNA
SWISS FEDERAL INSTITUTE OF TECHNOLOGY – LAUSANNE

Faculté Informatique et Communication
Introduction to Natural Language Processing (Ms; CS-431)
Chappelier, J.-C. & Rajman, M.

CS-431 Hands On Part-of-Speech tagging (part 2)

J.-C. Chappelier M. Rajman

v. 20251014 – 1

QUESTION I [2 pt]

(from Spring 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

Consider two sequences of discrete random variables $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$, with possibles values respectively $(x_1, x_2, \ldots)$ in V, $(y_1, y_2, \ldots)$ in T.

Indicate which of the following statements are always true (without any further assumption):

[✓] $$\sum_{(x_1,x_2,\ldots,x_n)\in V^n} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,|\, Y_1 = y_1, Y_s = y_2, \ldots, Y_n = y_n) = 1$$

$$\sum_x P(x|y) = 1$$

[ ] $$\sum_{(y_1,y_2,\ldots,y_n)\in T^n} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,|\, \underbrace{Y_1 = y_1, Y_s = y_2, \ldots, Y_n = y_n}_{\text{known}}) = 1$$

[✓] $P(Y_1, Y_2, \ldots, Y_n) = P(Y_n) \cdot P(Y_{n-1}|Y_n) \cdot \cdots \cdot P(Y_2|Y_3, \ldots, Y_n) \cdot P(Y_1|Y_2, \ldots, Y_n)$

[ ] $P(X_i|X_1, \ldots, X_{i-1}, Y_1, Y_2, \ldots, Y_n) = P(X_i|Y_i)$, for all i between 2 and n.

$$P(Y_1^n) = P(Y_{56}) \cdot P(Y_{13}|Y_{56}) \cdot P(Y_{27}|Y_{13}\, Y_{56}) \cdot P(Y_1|Y_{27}\, Y_{13}\, Y_{56}) \cdot \ldots \cdot P(Y_{42}|\overbrace{Y_1 \ldots Y_n}^{\text{but } Y_{42}})$$

continues on back ☞

QUESTION II [1 pt]

(from Spring 2018 quiz 2)

When using Hidden Markov Models to perform PoS tagging:

① What do the observables of the HMM model correspond to? → Words

② What do the hidden states of the HMM model correspond to?

→ tags

QUESTION III [2 pt]

(from Spring 2018 quiz 2)

For this question, *one or more* assertions can be correct. Tick only the correct assertion(s). There will be a penalty for wrong assertions ticked.

Indicate which of the following statements are true, when using Hidden Markov Models to perform PoS tagging:

- [ ] the Viterbi algorithm is used to efficiently ~~train~~ an HMM model on supervised data; Max likelihood
- [✓] the Baum-Welch algorithm can be used to efficiently train an HMM model on unsupervised data;
- [ ] provided that enough unsupervised data are available, the Baum-Welch algorithm is always able to learn the best possible HMM model;
- [ ] when an order-1 HMM is used, the assignment of a tag to a word only depends on the tag, the word, and the previous tag.

continues on back ☞

QUESTION IV [7 pt]

(from Fall 2018 quiz 2)Indicate the sequence of PoS tags assigned by an order-1 HMM to the word sequence "*iron shaped cloth*", if the following information is available:

Lexicon excerpt: (but no other tag for the provided words)

"*iron*": `Noun`, `Verb` "*shaped*": `Adj`, `Verb` "*cloth*": `Noun`

1- drawing (lattice with numbers)

init (**(some) Parameters:**

$P_I(\texttt{Noun}) = 2 \cdot 10^{-9}$ $P_I(\texttt{Verb}) = 1 \cdot 10^{-9}$ $P_I(\texttt{Adj}) = 3 \cdot 10^{-9}$

emit (

$P(\text{“}iron\text{''}|\texttt{Noun}) = 3 \cdot 10^{-9}$ $P(\text{“}shaped\text{''}|\texttt{Adj}) = 2 \cdot 10^{-9}$

$P(\text{“}iron\text{''}|\texttt{Verb}) = 2 \cdot 10^{-9}$ $P(\text{“}shaped\text{''}|\texttt{Verb}) = 3 \cdot 10^{-9}$

trans (

$P(\texttt{Adj}|\texttt{Noun}) = 1 \cdot 10^{-9}$ $P(\texttt{Adj}|\texttt{Verb}) = 2 \cdot 10^{-9}$

$P(\texttt{Verb}|\texttt{Noun}) = 2 \cdot 10^{-9}$ $P(\texttt{Verb}|\texttt{Verb}) = 1 \cdot 10^{-9}$

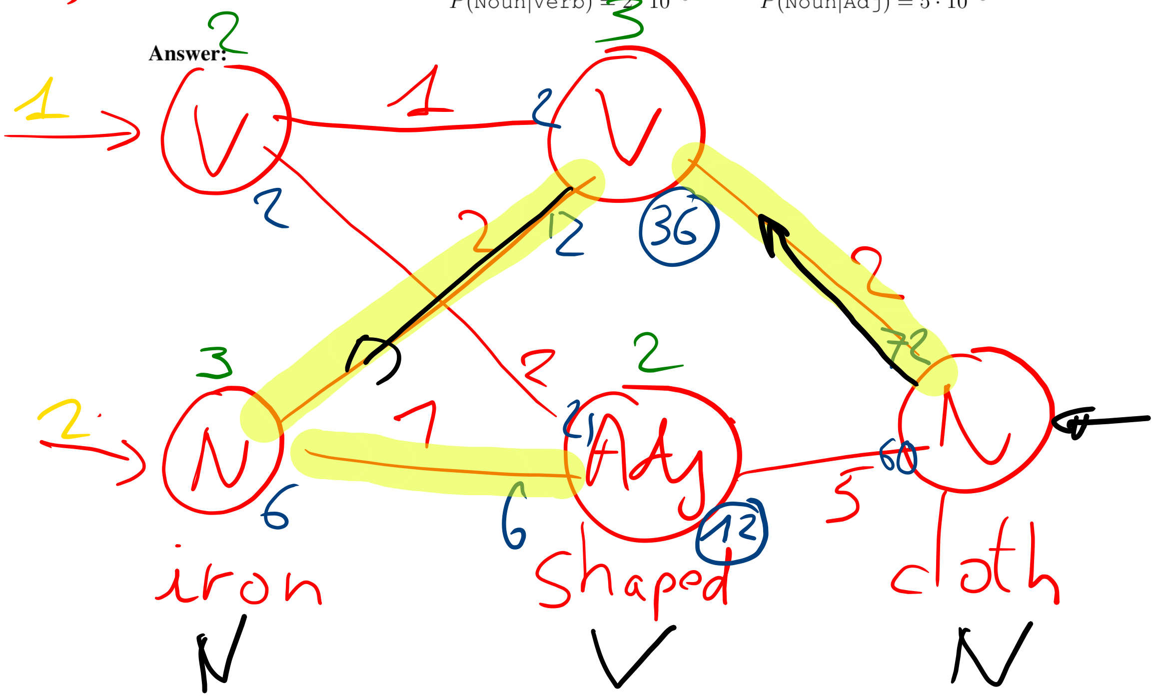
$P(\texttt{Noun}|\texttt{Verb}) = \cancel{2} \cdot 10^{-9}$ $P(\texttt{Noun}|\texttt{Adj}) = 5 \cdot 10^{-9}$

2- forward

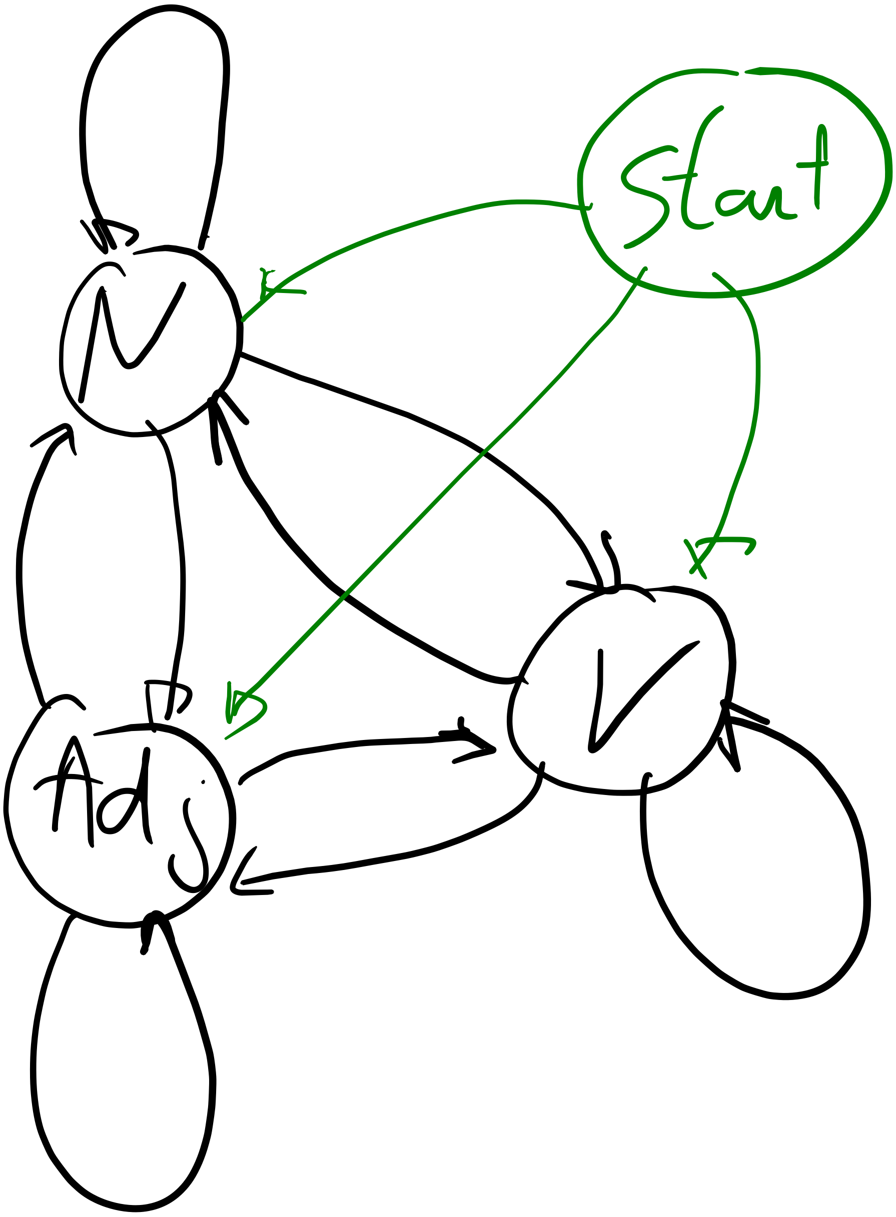
• product

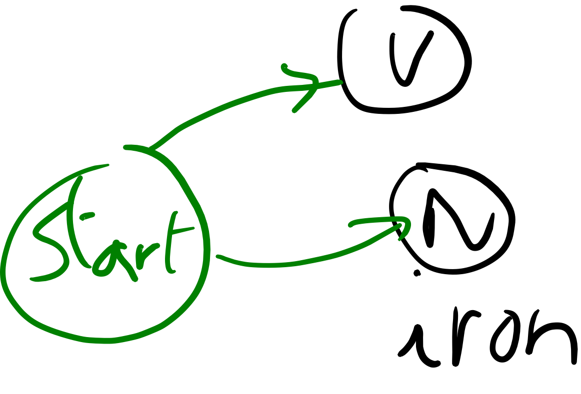
• max

Answer:



3- backward reconstruct max path


N
Start
Adj
V


V
Start
N
iron