Lecture reviews — Week 06

J.-C. Chappelier & M. Rajman

Laboratoire d'Intelligence Artificielle
Faculté I&C

Week(s 5 &) 6 keypoints

- HMM (technical)

↳ 3 problems

$$P(w_1^n|\theta) = \sum_{t_1^n} P(w_1^n t_1^n|\theta)$$

Viterbi ⟶ $\underset{t_1^n}{\text{Argmax}}\, P(t_1^n|w_1^n,\theta)$

$P(\theta|w_1^n)$ Learning (unsupervised)

($P(\theta|w_1^n,t_1^n)$)

$$P(W_1 = w_1, \ldots, W_n = w_n, T_1 = t_1, \ldots T_n = t_n)$$

$$P(\text{1st word} = \text{the}, \text{2 word} = \text{cat}, \ldots, \text{1 tag} = D, \ldots)$$

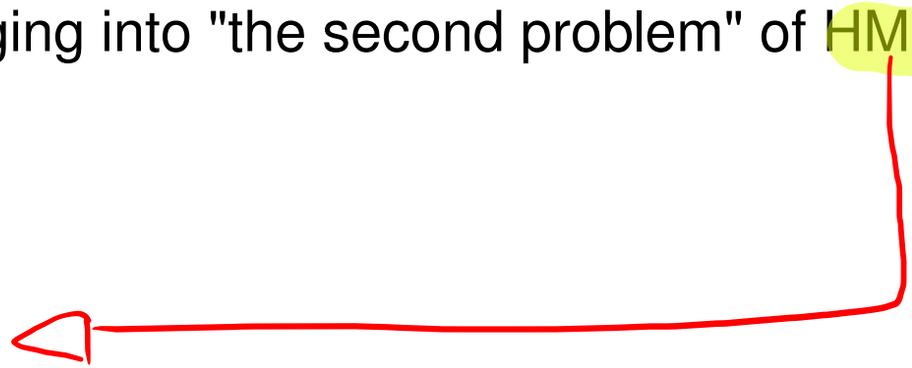
Week(s 5 &) 6 keypoints

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

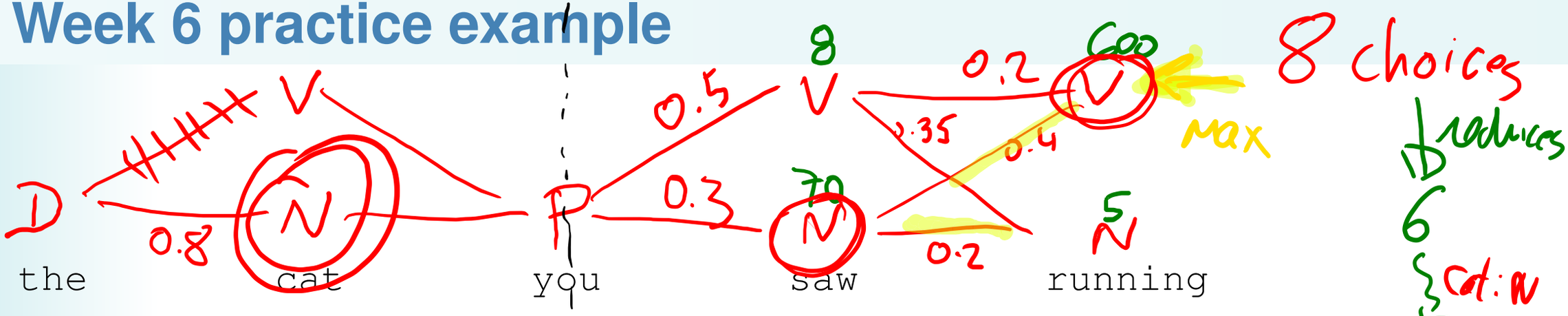
Week 6:

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



J.-C. Chappelier & M. Rajman

Week 6 practice example



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
```

```
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
```

1) formulas (8!!) → 4 (1) some cancels out

$P(D) \cdot P(\text{the}|D) \quad P(N|D) \cdot \ldots$

init emit transition

2) Viterbi

1) max incoming arcs

2) max at the end

3) reconstruct back