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. 20221026 – 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):

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

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

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

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

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:

wrong [ ] the Viterbi algorithm can be used to efficiently train an HMM model on supervised data;

[x] the Baum-Welch algorithm can be used to efficiently train an HMM model on unsupervised data;

wrong [ ] provided that enough unsupervised data are available, the Baum-Welch algorithm is always able to learn the best possible HMM model;

wrong [ ] 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.

Argmax $P(t_1)\,P(w_1|t_1)\cdot P(t_2|t_1)\cdots P(w_i|t_i)\,P(t_i|t_{i-1})\cdots$

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`

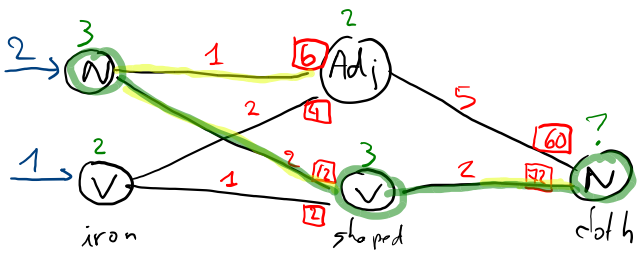
(some) Parameters:

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

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

$P(\texttt{Adj}|\texttt{Noun}) = 1 \cdot 10^{-9}$ $P(\texttt{Adj}|\texttt{Verb}) = 2 \cdot 10^{-9}$
$P(\texttt{Verb}|\texttt{Noun}) = 2 \cdot 10^{-9}$ $P(\texttt{Verb}|\texttt{Verb}) = 1 \cdot 10^{-9}$
$P(\texttt{Noun}|\texttt{Verb}) = 2 \cdot 10^{-9}$ $P(\texttt{Noun}|\texttt{Adj}) = 5 \cdot 10^{-9}$

Answer:



Order-2 HMM?

states: bigram of tags