PEE

Exercise 2

Jaime Ferrando Huertas

January 2020

1 Theoretical questions

1.1 Question 1

For the first question we need to reason the Inside-Outside behaviour when all initial probabilities for all rules are equal. If all rules have initial probabilities then Inside-Outside rules probability will be balance depending on how many times you can observe each rule in the samples we have. If we were to have equal initial probabilities in the example on page 50 from Part I.4 then rules like ($Nom \rightarrow$ "$vieja$") and ($Nom \rightarrow$ "$mujer$") will have equal probabilities as they appear once as a Nom. We think this initialization is good for cases when you don't have many samples from a language as it will not "overfit" to certain samples.

1.2 Question 3

For this question we are required to estimate the rule (Suj $\rightarrow$ Art Nom Adj) with the inside algorithm with an example of courses slides. We use the sample $\mathcal{D} = \{$*(la vieja)(demanda ayuda), la mujer oculta pelea, la vieja ayuda*$\}$. We compute it as follows:

$$P(Suj \rightarrow Art, Nom, Adj) = \frac{\frac{0.01176}{0.01266}}{\frac{0.0009}{0.0009} + \frac{0.0009+0.01176}{0.01266} + \frac{0.007}{0.007}} = 0.3096 \tag{1}$$

Used probabilities:

$$\begin{gathered} P_\theta((lavieja)(demandaayuda)) = 9 * 10^{-4} \\ P_\theta(lamujerocultapelea) = 0.01266 \\ P_\theta(laviejaayuda) = 0.007 \end{gathered} \tag{2}$$

1.3 Question 4

We are now required to estimate the same rule but with Viterbi algorithm. Is computed as follows:

$$P(Suj \to Art, Nom, Adj) = \frac{\frac{0.01176}{0.01176}}{\frac{0.0009}{0.0009} + \frac{0.0009+0.01176}{0.01266} + \frac{0.007}{0.007}} = 0.3333 \qquad (3)$$

Now is easier to calculate as Viterbi only takes the maximum probability tree while inside-outside takes into account all possible trees. Viterbi algorithm offers a good estimation (very close to result from question 3) while reducing compute time.

1.4 Question 5

This question ask us again to compute the rule (Suj → Art Nom Adj) with a new training sample $\mathcal{D}$ = {la vieja demanda ayuda, la mujer oculta pelea, la vieja mujer oculta demanda ayuda}. This training sample can not be accepted by the original rules and we are forced to introduce a new rule (Suj → Art , Adj, Nom, Adj) with probability 0.2, this means we also have to reduce the probability for other rules. Our final set of rules and their probability:

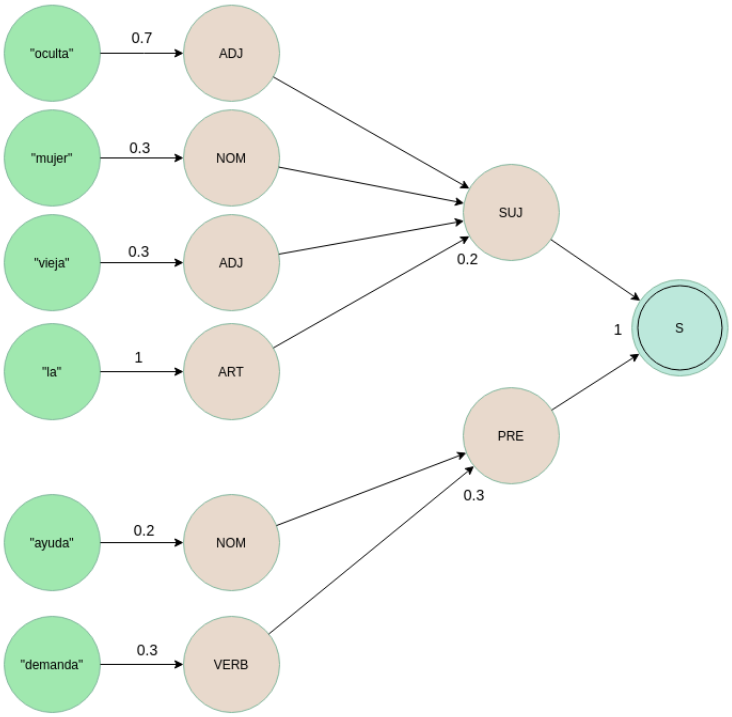
Rule	Probability
Suj → Art, Nom	0.4
Suj → Art, Adj, Nom	0.2
Suj → Art, Nom, Adj	0.2
Suj → Art, Adj, Nom, Adj	0.2

Table 1

Applying inside-outside algorithm we obtain the following estimation:

$$P(Suj \to Art, Nom, Adj) = \frac{\frac{0.01176}{0.01248}}{\frac{0.00072+0.00168}{0.0024} + \frac{0.00072+0.01176}{0.01248} + \frac{0.0002268}{0.0002268}} = 0.3141 \qquad (4)$$

Here is a plot for the sentence we just computed and its assigned probabilities.



1.5 Question 6

We now re-estimate the probability of question 5 with k-best and k=2. Viterbi is equivalent to k-best algorithm when k=1 but if k=(Number of tress) k-best will be equivalent to inside-outside. In our case we have 2 trees as seen in figure (4) so the result will be the same as in question 5.

$$P(Suj \rightarrow Art, Nom, Adj) = 0.3141 \tag{5}$$

2 Practical questions

2.1 Question 8

For this question we trained four different grammars with the Inside-Outside algorithm and trained them in the *SampleTriangle-10K* datasets. Each of the trained grammars has a different number of non-terminal symbols.

We first generated the grammars and then trained them for 700 iterations following the lab instructions. We then generated 1000 strings from each grammars and checked how many were rectangle, results presented.

N Terminals	N rectangles
5	29
10	63
15	61
20	84

Table 2: Number of rectangle triangles depending on number of terminals in the trained grammar

2.2 Question 9

For this last question we studied grammars trained with samples with and without brackets while testing a new Algorithm, Viterbi. We had to modify the training files in the DATA folder and remove brackets, then we just had to follow the same steps as previous question adding the -v instruction when Viterbi was being used. We computed their classification error with the provided confus script and present it on the following table.

Experiment	Error
Inside-Outside with brackets	41.13%
Inside-Outside without brackets	36.2%
Viterbi with brackets	58.7%
Viterbi without brackets	58.27%

Table 3: Classification error for Inside-Outside and Viterbi trained with and without brackets

We find that Inside-Outside stands as the best choice, specially when training with samples with removed brackets where it sees a higher improvement than Viterbi. This is not a surprise for us since Inside-outside computes the optimal solution while Viterbi is not guaranteed, despite this Viterbi can also be useful when training time is important and we now know removing brackets wont be a sustancial improvements.