STATISTICAL STRUCTURED PREDICTION

Question set (Part 1)

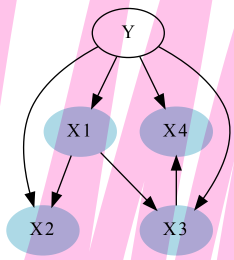
December 2023

Administrative issues

- This quiz represents the 30% of the final score of PEE.
- Each question has a score according to its difficulty.
- Answers can be handwritten, as long as this is readable.
- Please present the results in a readable "PDF" document and send it to me by email.
- Deadline for delivering these questions is **22 December 2023**

Question 1 (5 points)

Tree-augmented Naive Bayes classifier



Given the previous Bayesian Network that defines the tree-augmented Naive Bayes classifier, provide the resulting factorization for the joint probability $p(x_1^T, y)$.

Question 2 (10 points)

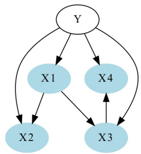
Briefly explain why languages generated by a probabilistic grammar are not necessarily probabilistic and why not all probabilistic languages can be generated by a probabilistic grammar.

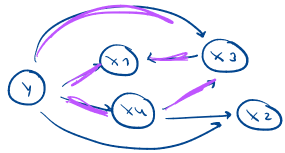
Question 3 (10 points)

Prove that the normalization factor of a CRF can be calculated as $Z(x;\theta) = \sum_s \alpha_T(s)$, with α_T being the *forward* score of the string $x_1 \dots x_T$.

Question 1 (5 points)

Tree-augmented Naive Bayes classifier





Given the previous Bayesian Network that defines the tree-augmented Naive Bayes classifier, provide the resulting factorization for the joint probability $p(x_1^T, y)$.

poner definiciones

$$p(y, x_1, x_4, x_2, x_3) = p(y) \cdot p(x_1 | y, x_3) \cdot p(x_4 | y) \cdot p(x_3 | y, x_4) \cdot p(x_2 | y, x_4)$$

Question 2 (10 points)

Briefly explain why languages generated by a probabilistic grammar are not necessarily probabilistic and why not all probabilistic languages can be generated by a probabilistic grammar.

Question 3 (10 points)

Prove that the normalization factor of a CRF can be calculated as $Z(x;\theta) = \sum_s \alpha_T(s)$, with α_T being the *forward* score of the string $x_1 \dots x_T$.

Where we can compute $Z(x;\ \theta)$ efficiently using the forward algorithm

$$Z(x;\ \theta) \ = \sum_{y'^T_1} \prod_{t=1}^{T} \Psi_t(y'_{t-1},\ y'_t,\ x_t) \ = \sum_s \alpha_T(s)$$

➢ **Definition**: Score for $x_1 \dots x_t$ ending at $y_t = s \in \mathcal{Y}$

$$\alpha_t(s) \ \stackrel{\text{def}}{=} \sum_{y_1^t;\ y_t = s} \prod_{i=1}^{t} \Psi_i\,(y_{i-1},\ y_i,\ x_i)$$

➢ **Initialization**: $\forall\, s \in \mathcal{Y}$

$$\alpha_1(s) \ = \ \Psi_1\,(y_0 = \text{null},\ y_1 = s,\ x_1)$$

➢ **Recursion**: $\forall\, t = 2 \dots T$; and $\forall\, s \in \mathcal{Y}$

$$\alpha_t(s) \ = \sum_{s' \in \mathcal{Y}} \alpha_{t-1}(s') \ \cdot \ \Psi_t\,(y_{t-1} = s',\ y_t = s,\ x_t)$$

➢ **Final result**: The optimal score for x is:

$$\sum_s \alpha_T\ (s)$$

3

$$Z(x;\theta) = \sum_{s} \alpha_T(s) = \sum_{y'^T_1} \prod_{t=1}^{T} \Psi_1(y'_{t-1}, y'_t, x_t)$$

$$\Psi_t(y_{t-1},\ y_t,\ x_t) \ \stackrel{\text{def}}{=}\ \exp\left\{\sum_{k=1}^{K} \theta_k\ f_k(y_{t-1},\ y_t,\ x_t)\right\}$$

- Demostración:

$$Z(x;\theta) = \Psi_1 \cdot (y_0 = \text{null}, y_1, x_1) + \alpha_1(y_1) \cdot \Psi_2(y_1, y_2, y_2) +$$

$$+ \alpha_2(y_2) \cdot \Psi_3(y_2, y_3, y_3) + \ldots + \alpha_{n-1}(y_{n-1}) \cdot \Psi_n(y_{n-1}, y_n, x_n) =$$

$$= \Psi_1 \cdot (y_0 = \text{null}, y_1, x_1) + \Psi_1 \cdot (y_0 = \text{null}, y_1, x_1) \cdot \Psi_2(y_1, y_2, y_2) +$$

$$+ \Psi_2(y_1, y_2, y_2) \cdot \Psi_3(y_2, y_3, y_3) + \ldots\ldots\ldots\ldots +$$

$$+ \Psi_{n-1}(y_{n-2}, y_{n-1}, x_{n-1}) \cdot \Psi_n(y_{n-1}, y_n, x_n) =$$

$$= \sum_{y'^T_1} \prod_{t=1}^{T} \Psi_t(y'_{t-1}, y_t, x_t) = \sum_{s} \alpha_T(s)$$

$$\sum_{y'^T_1} \prod_{t=1}^{T} \Psi_t(y'_{t-1},\ y'_t,\ x_t) = \sum_{s} \alpha_T(s)$$

3

$$Z(x;\theta) = \sum_s \alpha_T(s) = \sum_{y'^T_1} \prod_{t=1}^{T} \psi_t(y'_{t-1}, y'_t, x_t)$$

$$\Psi_t(y_{t-1},\ y_t,\ x_t) \overset{\text{def}}{=} \exp\left\{\sum_{k=1}^{K} \theta_k\ f_k(y_{t-1},\ y_t,\ x_t)\right\}$$

- Demostración:

$$Z(x;\theta) = \sum_{y'^T_1} \prod_{t=1}^{T} \psi_t(y'_{t-1}, y_t, x_t) =$$

$$= \sum_{y_T} \sum_{y_{T-1}} \Psi_T(y_T, y_{T-1}, x_T) \sum_{y_{T-2}} \Psi_{T-1}(y_{T-1}, y_{T-2}, x_{T-1}) \sum_{y_{T-3}} \cdots$$

$$\psi_1 \cdot (y_0 = \text{null}, y_1, x_1) + \alpha_1(y_1) \cdot \psi_2(y_1, y_2, y_2) +$$

$$+ \alpha_2(y_2) \cdot \psi_3(y_2, y_3, y_3) + \ldots + \alpha_{n-1}(y_{n-1}) \cdot \psi_n(y_{n-1}, y_n, x_n) =$$

$$= \psi_1 \cdot (y_0 = \text{null}, y_1, x_1) + \psi_1 \cdot (y_0 = \text{null}, y_1, x_1) \cdot \psi_2(y_1, y_2, y_2) +$$

$$+ \psi_2(y_1, y_2, y_2) \cdot \psi_3(y_2, y_3, y_3) + \ldots\ldots\ldots\ldots +$$

$$+ \psi_{n-1}(y_{n-2}, y_{n-1}, x_{n-1}) \cdot \psi_n(y_{n-1}, y_n, x_n) =$$

cómo desarrollo dps lo de $\sum, \prod$

= algo más ?? =

$$= \qquad = \sum_s \alpha_T(s)$$

$$\sum_{y'^T_1} \prod_{t=1}^{T} \Psi_t(y'_{t-1},\ y'_t,\ x_t) = \sum_s \alpha_T(s)$$

$$Z(x;\theta) = \sum_{s} \alpha_T(s) = \sum_{y'^T_1} \prod_{t=1}^{T} \Psi_1(y'_{t-1}, y'_t, x_t)$$

$$\Psi_t(y_{t-1}, y_t, x_t) \stackrel{\text{def}}{=} \exp\left\{\sum_{k=1}^{K} \theta_k f_k(y_{t-1}, y_t, x_t)\right\}$$

- Demostración:

$$Z(x;\theta) = \sum_{y'^T_1} \prod_{t=1}^{T} \Psi_1(y'_{t-1}, y'_t, x_t) = \sum_{y'^T_1} \prod_{t=1}^{T} \exp\left(\sum_{k=1}^{K} \theta_k \cdot f_k(y_{t-1}, y_t, x_t)\right) =$$

Proof:

The forward score $\alpha_t(y)$ is calculated using the forward algorithm:

$$\alpha_t(y) = \sum_{y' \in \mathcal{Y}} \alpha_{t-1}(y') \cdot e^{\sum_{t=1}^{T} \theta^T f(y'_{t-1}, y'_t, x, t)}$$

Now, let's consider the normalization factor $Z(x, \Theta)$:

$$Z(x, \Theta) = \sum_{y \in \mathcal{Y}} e^{\sum_{t=1}^{T} \alpha_t(y)}$$

Substitute the expression for $\alpha_t(y)$:

$$Z(x, \Theta) = \sum_{y \in \mathcal{Y}} e^{\sum_{t=1}^{T} \left(\sum_{y' \in \mathcal{Y}} \alpha_{t-1}(y') \cdot e^{\sum_{t=1}^{T} \theta^T f(y'_{t-1}, y'_t, x, t)} \right)}$$

Rearrange the terms:

$$Z(x, \Theta) = \sum_{y \in \mathcal{Y}} \sum_{y' \in \mathcal{Y}} e^{\sum_{t=1}^{T} \left(\alpha_{t-1}(y') + \theta^T f(y'_{t-1}, y'_t, x, t) \right)}$$

Now, recognize that the expression inside the sum is the forward score $\alpha_t(y')$ at position t for the label sequence y':

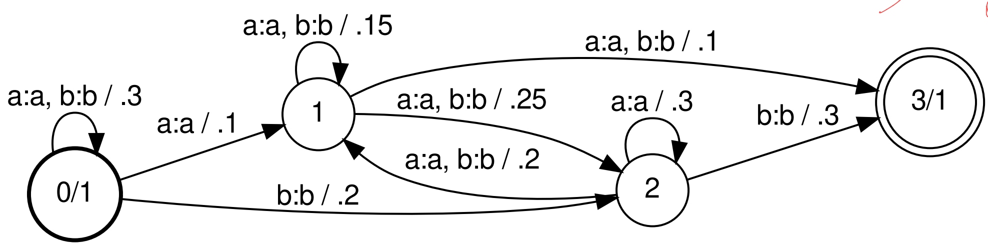
$$Z(x, \Theta) = \sum_{y \in \mathcal{Y}} \sum_{y' \in \mathcal{Y}} e^{\sum_{t=1}^{T} \alpha_t(y')}$$

Question 4 (30 points)

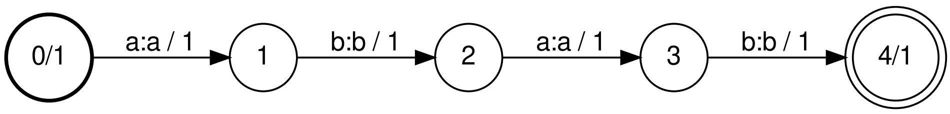
Given the following probabilistic transducers, T_1 and T_2, obtain the transducer resulting from the composition operation $(T_1 \circ T_2)$. The solution must leave a detailed record of all the steps to achieve it.

→ librería python OpenFST pywrapfst

T_1 :



T_2 :



Question 5 (30 points)

Suppose an instructor wants to determine whether a student has understood the class material based on the exam score. For this purpose, he has the following Bayesian network defined by the factorization associated with its joint probability.

CAMBIAR O!!

$$P(I, H, U, E) = P(I)\ P(H)\ P(U \mid I, H)\ P(E \mid U),$$

Where:

- I represents whether the student is very intelligent (yes or no)
- H represents whether the student is very hardworking (yes or no)
- U represents whether the student has understood the class material (yes or no)
- E represents whether the student has obtained a high score on the exam (yes or no)

The probabilities are,

$P(I)$

yes	0.3
no	0.7

$P(H)$

yes	0.4
no	0.6

$P(U \mid I, H)$

I	H	yes	no
yes	yes	0.7	0.3
yes	no	0.1	0.9
no	yes	0.1	
no	no	0.1	

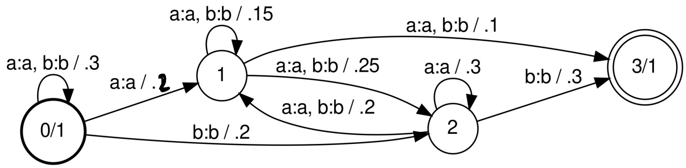
$P(E \mid U)$

U	yes	no
yes	0.9	0.1
no	0.2	0.8

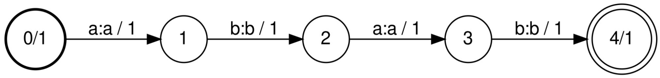
¿Qué es lo nuevo?

What is the probability that a student who obtained a high exam score understood the class material?

T_1 :



T_2 :



EST 0

T_1

0 a:a / 0,3 0

0 b:b / 0,3 0

0 a:a / 0,2 1

0 b:b / 0,2 1

T_2

0 a:a / 1 1 ⇒ (0,0) a:a / 0,2 (1,2)

⇒ 0 b:b / 0,2 2

EST 1

1 a:a / 0,15 1

1 b:b / 0,15 1

1 a:a / 0,1 3

1 b:b / 0,1 3

1 a:a / 0,1 2

1 b:b / 0,1 2

1 b:b / 1 2

⇒ 1 b:b / 0,15 2

2 a:a / 0,2 1

2 a:a / 0,3 2

2 b:b / 0,3 3

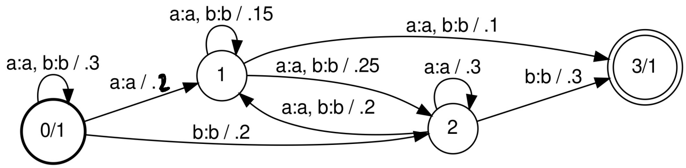
2 a:a / 1 3 ⇒ 1 a:a / 0,1 3

3 b:b / 1 4

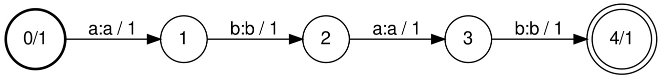
⇒ 1 b:b / 0,1 4

⇒ 2 b:b / 0,3 4

T_1 :



T_2 :



EST 0

T_1

0 a:a / 0,3 0

0 b:b / 0,3 0

0 a:a / 0,2 1

0 b:b / 0,2 1

T_2

0 a:a / 1 1 ⇒ (0,0) a:a / 0,3 (0,1)

⇒ (0,0) a:a / 0.2) (1,1)

EST 1

1 a:a / 0,15 1

1 b:b / 0,15 1

1 a:a / 0,1 3

1 b:b / 0,1 3

1 a:a / 0,1 2

1 b:b / 0,1 2

1 b:b / 1 2 ⇒ (0,1) b:b / 0,3 (0,1)

⇒ (0,1) b:b / 0,2 (1,1)

⇒ (1,1) b:b / 0,15 (1,2)

⇒ (1,1) b:b / 0,1 (3,2)

⇒ (1,1) b:b / 0,1 (2,2)

2 a:a / 0,2 1

2 a:a / 0,3 2

2 b:b / 0,3 3

2 a:a / 1 3

⇒ (0,2) a:a / 0,3 (0,3)

⇒ (0,2) a:a / 0,2 (1,3)

⇒ (1,2) a:a / 0,15 (1,3)

⇒ (1,2) a:a / 0,1 (3,3)

⇒ (1,2) a:a / 0,1 (2,3)

⇒ (2,2) a:a / 0,2 (1,3)

⇒ (2,2) a:a / 0,3 (2,3)

3 b:b / 1 4

⇒ (0,3) b:b / 0,3 (0,4)

⇒ (0,3) b:b / 0,2 (1,4)

⇒ (1,3) b:b / 0,15 (1,4)

⇒ (2,3) b:b / 0,1 (3,4)

⇒ (1,3) b:b / 0,1 (2,4)

⇒ (2,3) b:b / 0,3 (3,4)

$\Rightarrow$ (0,0) a:a / 0,3 (0,1) →

$\Rightarrow$ (0,0) a:a / 0.2) (1,1) →

$\Rightarrow$ (0,1) b:b / 0,3 (0,1) →

$\Rightarrow$ (0,1) b:b / 0,2 (1,1) →

$\Rightarrow$ (1,1) b:b / 0,15 (1,2)

$\Rightarrow$ (1,1) b:b / 0,1 (3,2)

$\Rightarrow$ (1,1) b:b / 0,1 (2,2)

$\Rightarrow$ (0,2) a:a / 0,3 (0,3)
$\Rightarrow$ (0,2) a:a / 0,2 (1,3)
$\Rightarrow$ (1,2) a:a / 0,15 (1,3)
$\Rightarrow$ (1,2) a:a / 0,1 (3,3)
$\Rightarrow$ (1,2) a:a / 0,1 (2,3)
$\Rightarrow$ (2,2) a:a / 0,2 (1,3)
$\Rightarrow$ (2,2) a:a / 0,3 (2,3)

$\Rightarrow$ (0,3) b:b / 0,3 (0,4)
$\Rightarrow$ (0,3) b:b / 0,2 (1,4)
$\Rightarrow$ (1,3) b:b / 0,15 (1,4)
$\Rightarrow$ (1,3) b:b / 0,1 (3,4)
$\Rightarrow$ (1,3) b:b / 0,1 (2,4)
$\Rightarrow$ (2,3) b:b / 0,3 (3,4)

5

$$P(U|E) = \frac{P(U=yes, E=yes)}{P(E=yes)}$$

incondicionalmente independ.

- Calculamos $P(U=yes, E=yes)$

$$P(U=yes, E=yes) = \sum_{I}\sum_{H} P(I, H, U=yes, E=yes) =$$

$$= P(I=yes)\cdot P(H=yes)\cdot P(U=yes \mid I=yes, H=yes)\cdot P(E=yes \mid Y=yes) +$$

$$+ P(I=no)\cdot P(H=yes)\cdot P(U=yes \mid I=no, H=yes)\cdot P(E=yes \mid Y=yes) +$$

$$+ P(I=yes)\cdot P(H=no)\cdot P(U=yes \mid I=yes, H=no)\cdot P(E=yes \mid Y=yes) +$$

$$+ P(I=no)\cdot P(H=no)\cdot P(U=yes \mid I=no, H=no)\cdot P(E=yes \mid Y=yes) =$$

$$= 0{,}3\cdot 0{,}4\cdot 0{,}9\cdot 0{,}9 + 0{,}7\cdot 0{,}4\cdot 0{,}5\cdot 0{,}9 +$$

$$+ 0{,}3\cdot 0{,}6\cdot 0{,}6\cdot 0{,}9 + 0{,}7\cdot 0{,}6\cdot 0{,}1\cdot 0{,}9 =$$

$$= 0{,}0972 + 0{,}126 + 0{,}0972 + 0{,}0378 = \boxed{0{,}3582}$$

→

- Calculamos $P(E=yes)$ inncondicionalmente independ.

$$P(E=yes) = \sum_I \sum_H \sum_U P(I) \cdot P(H) \cdot P(U \mid I,H) \cdot P(E=yes \mid U) =$$

$$= P(I=yes) \cdot P(H=yes) \cdot P(U=yes \mid I=yes, H=yes) \cdot P(E=yes \mid U=yes) +$$
$$+ P(I=no) \cdot P(H=yes) \cdot P(U=yes \mid I=no, H=yes) \cdot P(E=yes \mid U=yes) +$$
$$+ P(I=yes) \cdot P(H=no) \cdot P(U=yes \mid I=yes, H=no) \cdot P(E=yes \mid U=yes) +$$
$$+ P(I=no) \cdot P(H=no) \cdot P(U=yes \mid I=no, H=no) \cdot P(E=yes \mid U=yes) +$$
$$+ P(I=yes) \cdot P(H=yes) \cdot P(U=no \mid I=yes, H=yes) \cdot P(E=yes \mid U=no) +$$
$$+ P(I=no) \cdot P(H=yes) \cdot P(U=no \mid I=no, H=yes) \cdot P(E=yes \mid U=no) +$$
$$+ P(I=yes) \cdot P(H=no) \cdot P(U=no \mid I=yes, H=no) \cdot P(E=yes \mid U=no) +$$
$$+ P(I=no) \cdot P(H=no) \cdot P(U=no \mid I=no, H=no) \cdot P(E=yes \mid U=no) =$$

$$= 0.3 \cdot 0,4 \cdot 0,9 \cdot 0,9 + 0,7 \cdot 0,4 \cdot 0,5 \cdot 0,9 + 0,3 \cdot 0,6 \cdot 0,6 \cdot 0,9 +$$
$$+ 0,7 \cdot 0,6 \cdot 0,1 \cdot 0,9 + 0,3 \cdot 0,4 \cdot 0,9 \cdot 0,2 + 0,7 \cdot 0,4 \cdot 0,5 \cdot 0,2 +$$
$$+ 0,3 \cdot 0,6 \cdot 0,6 \cdot 0,2 + 0,7 \cdot 0,6 \cdot 0,1 \cdot 0,2 =$$

$$= 0.0972 + 0,126 + 0,0972 + 0,0378 + 0,0216 + 0,028 +$$
$$+ 0.0216 + 0,0084 = 0,4378$$

- Una vez calculados $P(E=yes, U=yes)$ y $P(E=yes)$ calculamos $P(E=yes \mid U=yes)$

$$P(U \mid E) = \frac{P(U=yes, E=yes)}{P(E=yes)} = \frac{0,3582}{0,4378} = \boxed{0,8181}$$

El 81,81 % de los estudiantes que han obtenido una buena nota han entendido el material de clase.

Question 5 (30 points)

Suppose an instructor wants to determine whether a student has understood the class material based on the exam score. For this purpose, he has the following Bayesian network defined by the factorization associated with its joint probability.

$$P(I, H, U, E) = P(I)\, P(H)\, P(U \mid I, H)\, P(E \mid U)\,,$$

Where:

I	represents whether the student is very intelligent (yes or no)
H	represents whether the student is very hardworking (yes or no)
U	represents whether the student has understood the class material (yes or no)
E	represents whether the student has obtained a high score on the exam (yes or no)

The probabilities are,

$P(I)$

yes	0.3
no	0.7

$P(H)$

yes	0.4
no	0.6

$P(U \mid I, H)$

I	H	yes	no
yes	yes	0.7	0.3
yes	no	0.1	0.9
no	yes	0.1	
no	no	0.1	

$P(E \mid U)$

U	yes	no
yes	0.9	0.1
no	0.2	0.8

What is the probability that a student who obtained a high exam score understood the class material?

$(U \mid E)$

BNs: a detailed example

prob lluvia

$P(S \mid R)$

Rain (R)	Sprinkler (S) s	r
n	0.60	0.40
y	0.99	0.01

Aspersor depende solo lluvia

$P(R)$

Rain (R) n (o)	y (en)
0.8	0.2

Sprinkler	s:	stoped
	r:	running
Grass	d:	dry
	w:	wet
Rain	n:	not rain
	y:	yes it rains

$P(G \mid S, R)$

Rain (R)	Sprinkler (S)	Grass (G) d	w
n	s	0.99	0.01
n	r	0.10	0.90
y	s	0.20	0.80
y	r	0.01	0.99

el cesped puede estar mojado o seco y puede llover o ?

fact. de la prob conjunta = coincide con grafo

estado inicic.

Joint distribution: $P(R, S, G) \;=\; P(R)\;\; P(S \mid R)\;\; P(G \mid R, S)$

Exercise: What is the probability that the sprinkler will work if there is no rain and the grass is wet? → prob de q esté marcha de q ? llueve y césped moj lo demás es en t2 → mirar paper

Question 6 (30 points)

As introduced in class, a CRF can be viewed as a factor graph where the bigram transition scores can be expressed as:

$$\Psi_t\,(y_{t-1},\ y_t,\ x_t) = \exp\,\{\sum_{k=1}^{K}\ \theta_k\ f_k\,(y_{t-1},\ y_t,\ x_t)\,\}$$

We also saw that the Forward algorithm for CRFs could be used to calculate the score of an observation, $x_1 \ldots x_t$ ending at $y_t = s \in \mathcal{Y}$ as,

$$\alpha_t(s) \quad \stackrel{\text{def}}{=} \quad \sum_{y_1^t;\ y_t=s} \ \prod_{i=1}^{t} \ \Psi_i\,(y_{i-1},\ y_i,\ x_i)$$

In this question, we will consider a *"trigram"* model for CRF, where the transition, $\Psi_t(\,\cdot\,)$, takes the form,

$$\Psi_t\,(y_{t-2}, y_{t-1},\ y_t,\ x_t) \ = \ \exp\,\{\sum_{k=1}^{K} \theta_k\ f_k\,(y_{t-2}, y_{t-1},\ y_t,\ x_t\,\}\,.$$

$\Psi_t(\,\cdot\,)$ is again a local transition score. We have assumed in this definition that $y_{-1} = y_0 = \#$. We call it a trigram CRF because we now consider sequences of three output labels (y_{t-2}, y_{t-1}, y_t).

Taking as reference the Forward algorithm seen in class, give a new version of this Forward algorithm for a *trigram* CRF.

Question 7 (35 points) → buscar el inside

Given the CKY-based *Inside* algorithm shown in class. This algorithm calculates the probability of an input sequence, $x_1, x_2, \ldots, x_T$. The initialization step is as follows: $\forall A \in N$ and $\forall i = 0 \ldots T-1$,

$$e(A < i, i+1 >) \ = \ p(A \to b)\ \delta(b, x_{i+1})\,,$$

and the recursion step is as follows: $\forall A \in N$ and $l : 2 \ldots T$, $\forall i : 0 \ldots T-l$,

$$e(A < i, i{+}l >) \ = \ \sum_{B,C \in N} p(A \to BC) \sum_{k=1,\ldots,l-1} e(B < i, i{+}k >)\, e(C < i{+}k, i{+}l >)\,.$$

Finally, we return

$$P_{G_\theta}(x) = e(S < 0, n >)\,.$$

Now, assume that we want to compute the probability of an input sequence for a right-branching grammar. Here are some examples of right-branching trees:

6

- **Definición:** Score para $x_1 \dots x_t$ acabando a $y_t = s \in \mathcal{Y}$

$$\alpha_t(s) \overset{def}{=} \sum_{y_1^t;\, y_t = s} \prod_{i=1}^{t} \Psi_i(y_{i-2}, y_{i-1}, y_i, x_i)$$

- **Inicialización:** $\forall s \in \mathcal{Y}$

$$\alpha_1(y_1, y_2) = \Psi_1(y_0 = null,\ y_1 = null,\ y_2 = s,\ x_1)$$

$$\alpha_2(y_2, y_3) = \alpha_1(y_1, y_2) \cdot \Psi_2(y_0 = null,\ y_1 = s',\ y_2 = s)$$

- **Recursión:** $\forall t = 3 \dots T$

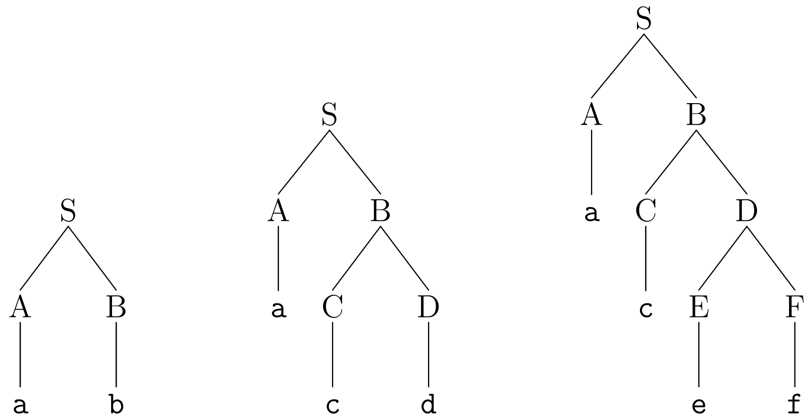
$$\alpha_t(y_t, y_{t-1}) = \sum_{s'' \in \mathcal{Y}} \cdot \sum_{s' \in \mathcal{Y}} \alpha_{t-1}(y_{t-2}, y_{t-1}) \cdot \Psi_t(y_{t-2} = s'',\ y_{t-1} = s',\ y_t = s,\ x_i)$$

- **Finalización:** el score óptimo para x es:

$$\sum_{s'' \in \mathcal{Y}} \cdot \sum_{s' \in \mathcal{Y}} \alpha_T(y_{T-1}, y_T)$$

Para esta formulación:

- $\alpha_t(y_t, y_{t-1})$ es el forward score en la posición t para la combinación de estados y_t, y_{t-1}.



As can be seen in the examples, the right-branching parse trees involve binary rules of form $(X \to Z\,Y)$, where the non-terminal, Z, must derive directly from a terminal symbol.

Give a new version of the CKY-based *Inside* algorithm to calculate the probability of an input sequence, $x_1, \ldots, x_T$, for (*right-branching*) grammars. In addition, the temporary cost of the algorithm proposed must also be determined.

Ejercicios de Repaso ? → transd.
→ demostrar. me he mirado reps pero no sé

➢ **Definition**: Given $x = x_1 \ldots x_T \in \Sigma^*$ and $A \in N$

$$e(A, i, i+l) \quad \stackrel{\text{def}}{=} \quad P_\theta(A \stackrel{*}{\Rightarrow} x_{i+1} \ldots x_{i+l})$$

➢ **Initialization**: $\forall\, A \in N; \qquad \forall\, i : 0 \ldots T-1;$

$$e(A, i, i+1) \;=\; p(A \to b) \;\cdot\; \delta(b, x_{i+1})$$

➢ **Recursion**: $\forall\, A \in N; \qquad \forall\, l : 2 \ldots T; \qquad \forall\, i : 0 \ldots T-l;$

quito B pq B es izq.

$$e(A, i, i+l) \;=\; \sum_{\cancel{B}, C \in N} \Big\{ p(A \to \cancel{B} C) \cdot$$

$$\sum_{k=1,\ldots,l-1} \cancel{e(B, i, i+k)} \;\cdot\; e(C, i+k, i+l) \Big\}$$

➢ **Final result**: The sentence probability is:

$$P_\theta(x) \;=\; e(S, 0, T)$$

muy simple?