|  |
| --- |
| Métodos probabilistas |
| Práctica del tema 1: Fundamentos de redes bayesianas |
|  |
| Gabriel Vázquez Torres |
|  |

Métodos probabilistas

Práctica del tema 1: Fundamentos de redes bayesianas

Contenido

[Descripción de la práctica 2](#_Toc502696683)

[Descripción de los corpus 3](#_Toc502696684)

[Brown Corpus 3](#_Toc502696685)

[Sussane Corpus 4](#_Toc502696686)

[Penn Treebank 5](#_Toc502696687)

[Comparativa de distintos aspectos 6](#_Toc502696688)

[Tipo de etiquetado 6](#_Toc502696689)

[Tamaño del corpus 6](#_Toc502696690)

[Tamaño del conjunto de etiquetas 6](#_Toc502696691)

[Temáticas incluidas 7](#_Toc502696692)

[Procedencia 8](#_Toc502696693)

[Breve análisis de qué corpus es el más apropiado diferenciando entre el corps de Brown y el de Sussane. 9](#_Toc502696694)

[Bibliografía 10](#_Toc502696695)

# Descripción de la práctica

Realizar diferentes ejercicios que se mostrarán en los siguientes puntos.

NOTA: Comentar que esta práctica la he realizado con WORD por falta de tiempo y por desconocimiento de Latex y Lyx. En las siguientes prácticas trataré de usar Latex.

# Ejercicio 1.1

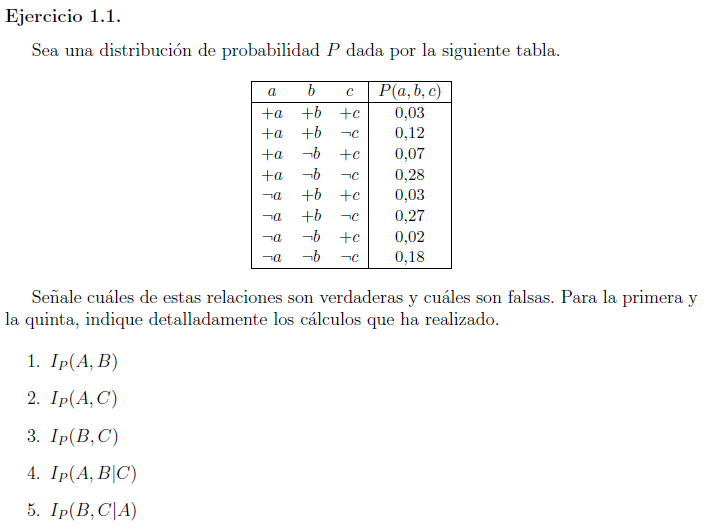


Ilustración 1: Ejercicio 1

En primer lugar vamos a calcular las probabilidades marginales de las siguientes combinaciones:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| -------------- | bc | ¬bc | ¬b¬c | b¬c | P(xi) |
| +a | 0,03 | 0,07 | 0,28 | 0,12 | 0,5 |
| ¬a | 0,03 | 0,02 | 0,18 | 0,27 | 0,5 |
| P(x,y) | 0,06 | 0,09 | 0,46 | 0,39 | 1,00 |

Tabla 1:Probabilidad marginal para "a"

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| -------------- | ac | ¬ac | ¬a¬c | a¬c | P(xi) |
| +b | 0,03 | 0,03 | 0,27 | 0,12 | 0,45 |
| ¬b | 0,07 | 0,02 | 0,18 | 0,28 | 0,55 |
| P(x,y) | 0,10 | 0,05 | 0,45 | 0,40 | 1,00 |

Tabla 2:Probabilidad marginal para "b"

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| -------------- | ab | ¬ab | ¬a¬b | a¬b | P(xi) |
| +c | 0,03 | 0,03 | 0,02 | 0,07 | 0,45 |
| ¬c | 0,12 | 0,27 | 0,18 | 0,28 | 0,55 |
| P(x,y) | 0,15 | 0,30 | 0,20 | 0,35 | 1,00 |

Tabla 3: Probabilidad marginal para "c"

Para todos los casos que no son condicionales, tenemos que:



Ilustración 2: Cálculo de la independencia probabilística

Para 1:

Como para que dos características sean independientes entre sí hay que realizar la operación anterior para todas las variables, hemos encontrado una en la que no se verifica la igualdad:

R = P(+a)=0,5

M = P(+b)=0,45

R\*M = 0,225

P(+a, +b) = 0,15

Como 0,15 no es igual a 0,225, la relación “1.” es Falsa por la fórmula anterior.

Para 2: Es Falsa.

Para 3. Es Falsa.

Teniendo en cuenta para la 4 y 5:

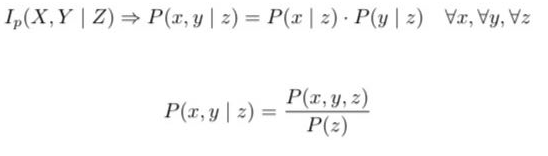


Ilustración 3: Cálculo de independencia probabilística con condición

Para 4 es Falsa

Para 5:

Teniendo en cuenta las fórmulas, cogemos un valor para A,B y C. En este caso: +a,+b,+c

En primer lugar calculamos P(+b,+c | +a) = P(+b,+c,+a)/P(+a) ->

(0,45 \* 0,15 \* 0,5)/0,5 = 0,0675

En segundo lugar calculamos P(+b|+a) = P(+b,+a)/P(+a) ->

0,15/0,5 = 0,3

En tercer lugar, calculamos: P(+c|+a) = P(+c,+a)/P(+a) ->

0,1/0,5 = 0,2

En cuarto lugar, comprobamos la igualdad:

0,0675 = 0,3 \* 0,2 🡪 0,0675 = 0,06

Ergo, después de realizar estos cálculos, podemos decir que la relación 5 es Falsa al no verificarse la igualdad.

# Ejercicio 1.2

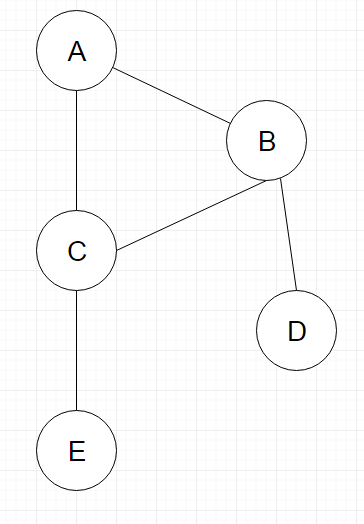


Ilustración 4: Representación del grafo del ejercicio 1.2

## Apartado 1:

Tenemos en cuenta:

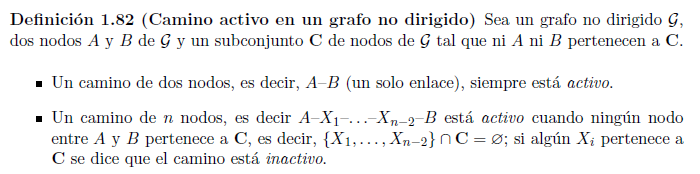


Ilustración 5: Definición de camino activo en grafo no dirigido.

Tenemos en cuenta la definición para grafo conectado y separado:

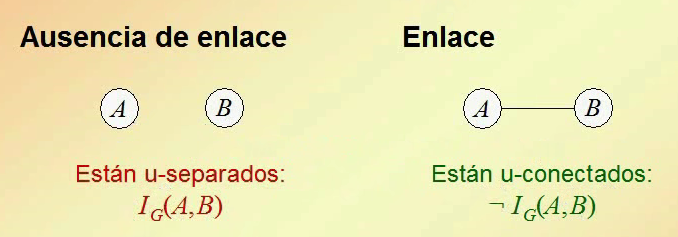


Ilustración 6: Representación de grafo separado o conectado (imagen obtenida del video “separación de grafos”).

Tenemos en cuenta que nos están pidiendo en todos los apartados, al no aparecer el signo “¬” delante de la “I” que significa “separados” o “independientes”, la AFIRMACIÓN de la separación entre sendos nodos de cada apartado.

1. Es FALSA ya que existen dos caminos conectados entre a y b: {A->B}, {A->C->B}
2. Es FALSA ya que existe un camino conectado entre b y d: {B->D}
3. Es FALSA ya que existen dos caminos conectados entre a y d: {A->B->D}, {A->C->B->D}
4. Es FALSA ya que existen dos caminos conectados entre d y e: {D->B->C->E}, { D->B->A->C->E }

## Apartado 2

Teniendo en cuenta la definición:

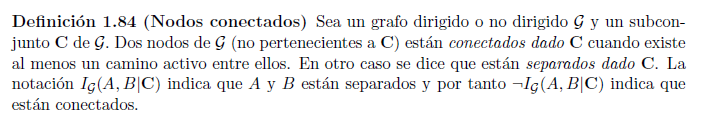
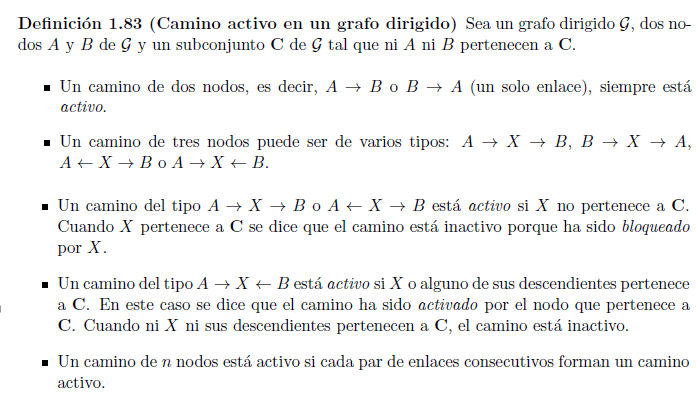


Ilustración 7: Definición de nodos conectados

Y la Ilustración 5:

1. Es FALSA ya que existe un camino posible entre a y b: {A->B}
2. Es VERDADERA ya que no existe ningún camino posible si el nodo B es el condicionante entre a y d.
3. Es VERDADERA ya que no existe ningún camino entre d y e si el nodo C es el condicionante.
4. Es FALSA ya que existe un camino entre d y e: { D->B->C->E}

# Ejercicio 1.3



# Bibliografía

|  |  |
| --- | --- |
| [1] | G. Rojo. [En línea]. Available: https://gramatica.usc.es/~grojo/Publicaciones/Lgca\_corpus\_lgca\_espanol.pdf. |
| [2] | F. y. Kucera, «aclweb,» [En línea]. Available: http://www.aclweb.org/anthology/C80-1006. |
| [3] | A. Lindebjerg. [En línea]. Available: http://www.hit.uib.no/icame/brown/bcm.html. |
| [4] | O. KHOKOVSKAIA. [En línea]. Available: http://radio.feld.cvut.cz/conf/poster/proceedings/Poster\_2017/Section\_HS/HS\_018\_Kholkovskaia.pdf. |
| [5] | [En línea]. Available: https://en.wikipedia.org/wiki/Brown\_Corpus. |
| [6] | [En línea]. Available: https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus\_ling/content/corpora/list/public/susanne.html. |
| [7] | [En línea]. Available: https://www.grsampson.net/SueDoc.html. |
| [8] | [En línea]. Available: https://research.csc.fi/-/susanne-corpus. |
| [9] | [En línea]. Available: http://www.lllf.uam.es/~sandoval/UAMTreebank.html. |
| [10] | [En línea]. Available: http://clic.ub.edu/. |
| [11] | [En línea]. Available: https://es.wikipedia.org/wiki/TreeBank. |
| [12] | «books.google,» [En línea]. Available: https://books.google.es/books?id=r3xyBgAAQBAJ&pg=PA238&lpg=PA238&dq=thematics+penn+treebank&source=bl&ots=rI5UVySG3q&sig=gVXpUTbH7dacg8NVA52idx\_LtHM&hl=en&sa=X&ved=0ahUKEwiPuaaB5rnYAhWKwbwKHWy1AqQQ6AEIMDAB#v=onepage&q=thematics%20penn%20treebank&f=false. |
| [13] | [En línea]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.8216&rep=rep1&type=pdf. |
| [14] | [En línea]. Available: https://www.grsampson.net/SueDoc.html. |
| [15] | [En línea]. Available: https://www.ling.upenn.edu/courses/Fall\_2003/ling001/penn\_treebank\_pos.html. |
| [16] | [En línea]. Available: https://www.clips.uantwerpen.be/pages/mbsp-tags. |
| [17] | [En línea]. Available: https://catalog.ldc.upenn.edu/docs/LDC95T7/cl93.html. |
| [18] | Ron Kohavi, George H. John, «Wrappers for feature subset selection,» Mountain View, 1995. |
| [19] | José Hernández Orallo, M. José Ramírez Quintana César, Introducción a la Minería de Datos. |