

Sesión 10: Redes Bayesianas – Aprendizaje

“Preferiría descubrir una ley causal
que ser rey de Persia” [Democritus]

Aprendizaje de Redes Bayesianas

- Introducción
- Aprendizaje paramétrico
 - Incertidumbre en las probabilidades
 - Datos faltantes / nodos ocultos
- Aprendizaje estructural
 - Árboles
 - Poliárboles
 - Redes multiconectadas
- Combinación de conocimiento y datos

Incertidumbre - RB III, L.E. Sucar

2

Aprendizaje

El aprendizaje inductivo consiste en
obtener el conocimiento a partir de datos.

En redes bayesianas se divide en 2 aspectos:

- Obtener la estructura de la red - estructural
- Obtener las probabilidades asociadas - paramétrico

Incertidumbre - RB III, L.E. Sucar

3

Aprendizaje Paramétrico

- Datos completos - se estiman las probabilidades a partir de frecuencias

$$P(A) \sim N_a / N_t$$

$$P(B|A_1, \dots, A_n) \sim N_{a_1, \dots, a_n, b} / N_{a_1, \dots, a_n}$$

Incertidumbre - RB III, L.E. Sucar

4

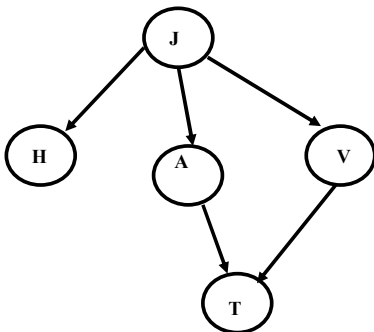
Ejemplo - ¿Cuándo jugar golf?

Ambiente	Temp.	Humedad	Viento	Jugar
soleado	alta	alta	no	N
soleado	alta	alta	si	N
nublado	alta	alta	no	P
lluvia	media	alta	no	P
lluvia	baja	normal	no	P
lluvia	baja	normal	si	N
nublado	baja	normal	si	P
soleado	media	alta	no	N
soleado	baja	normal	no	P
lluvia	media	normal	no	P
soleado	media	normal	si	P
nublado	media	alta	si	P
nublado	alta	normal	no	P
lluvia	media	alta	si	N

Incertidumbre - RB III, L.E. Sucar

5

Ejemplo – estructura



Incertidumbre - RB III, L.E. Sucar

6

Ejemplo

- $P(J)$
 - $P(N) = 5/14$
 - $P(P) = 9/14$
- $P(V|J)$
 - $P(si|N)=3/5$, $P(si|P)=3/9$
 - $P(no|N)=2/5$, $P(no|P)=6/9$
- Etc.

Incertidumbre - RB III, L.E. Sucar

7

Incertidumbre en las probabilidades

- Normalmente hay incertidumbre en las probabilidades, ya sea estimadas de datos o por expertos
- Se puede representar mediante una distribución de probabilidad, ventajas:
 - Representación explícita
 - Combinar información de expertos con datos
 - Propagar la incertidumbre en las probabilidades

Incertidumbre - RB III, L.E. Sucar

8

Incertidumbre en las probabilidades

- Variables binarias – distribución Beta

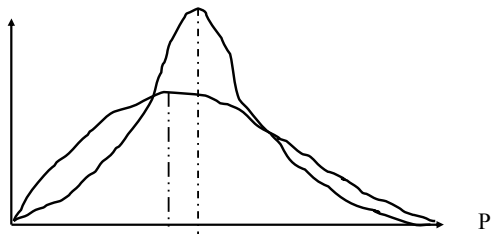
$$\beta(a, b) = \frac{(a+b+1)!}{a!b!} x^a (1-x)^b$$

- Valor promedio (esperado):
 $P(b1) = a+1 / a+b+2$

Incertidumbre - RB III, L.E. Sucar

9

Distribución Beta



Incertidumbre - RB III, L.E. Sucar

10

Incertidumbre en las probabilidades

- Modelación de estimación de expertos mediante valores de a y b :
 - Ignorancia completa: $a=b=0$
 - Poco confidente: $a+b$ pequeño (10)
 - Medianamente confidente: $a+b$ mediano (100)
 - Muy confidente: $a+b$ grande (1000)
- Combinación de experto y datos:
 - $P(b_1) = k+a+1 / n+a+b+2$
 - Datos: k/n
 - Experto: $a/a+b$

Incertidumbre - RB III, L.E. Sucar

11

Incertidumbre en las probabilidades

- Variables multivaluadas – se utiliza la generalización de la Beta $\rightarrow$ distribución *Dirichlet*

$$\text{Dir}(a_1, a_2, \dots, a_n) = \frac{(b_i + a_i + t - 1)!}{a_i! (b_i + t - 2)!} x^{a_i} (1-x)^{(b_i+t-1)}$$

– Donde: $b_i = [\sum_j a_j] - a_i$; $t = \#$ de valores

- Valor esperado:

$$a_i + 1 / a_1 + a_2 + \dots + a_t + t$$

Incertidumbre - RB III, L.E. Sucar

12

Información incompleta

- En la práctica, en muchas ocasiones los datos no están completos
- Dos tipos básicos de información incompleta:
 - Faltan algunos valores de una de las variables en algunos casos – datos incompletos
 - Faltan todos los valores de una variable – nodos ocultos

Incertidumbre - RB III, L.E. Sucar

13

Información incompleta

Ambiente	Temp.	Humedad	Viento	Jugar
soleado	xxx	alta	--	N
soleado	alta	alta	--	N
nublado	alta	alta	--	P
lluvia	media	alta	--	P
lluvia	baja	normal	--	P
lluvia	baja	normal	--	N
nublado	baja	normal	--	P
soleado	media	alta	--	N
soleado	xxx	normal	--	P
lluvia	media	normal	--	P
soleado	media	normal	--	P
nublado	media	alta	--	P
nublado	alta	normal	--	P
lluvia	media	alta	--	N

Incertidumbre - RB III, L.E. Sucar

14

Datos incompletos

Existen varias alternativas:

1. Considerar un nuevo valor “desconocido”
2. Tomar el valor más probable (promedio) de la variable
3. Considerar el valor más probable en base a las otras variables
4. Considerar la probabilidad de los diferentes valores en base a las otras variables

Incertidumbre - RB III, L.E. Sucar

15

Datos incompletos

Valor más probable:

1. Asignar todas las variables observables.
2. Propagar su efecto y obtener las probabilidades posteriores de las no observables.
3. Para las variables no observables, asumir el valor con probabilidad mayor como observado.
4. Actualizar las probabilidades previas y condicionales de acuerdo a las fórmulas anteriores.
5. Repetir 1 a 4 para cada observación.

Incertidumbre - RB III, L.E. Suar

16

Datos incompletos

Ambiente	Temp.	Humedad	Viento	Jugar
soleado	xxx	alta	--	N
soleado	alta	alta	--	N
nublado	alta	alta	--	P
lluvia	media	alta	--	P
lluvia	baja	normal	--	P
lluvia	baja	normal	--	N
nublado	baja	normal	--	P
soleado	media	alta	--	N
soleado	xxx	normal	--	P
lluvia	media	normal	--	P
soleado	media	normal	--	P
nublado	media	alta	--	P
nublado	alta	normal	--	P
lluvia	media	alta	--	N

$P(T|sol,alta,N)$

$P(T|sol,nor,P)$

Incertidumbre - RB III, L.E. Suar

17

Datos incompletos

Ambiente	Temp.	Humedad	Viento	Jugar
soleado	media	alta	--	N
soleado	alta	alta	--	N
nublado	alta	alta	--	P
lluvia	media	alta	--	P
lluvia	baja	normal	--	P
lluvia	baja	normal	--	N
nublado	baja	normal	--	P
soleado	media	alta	--	N
soleado	media	normal	--	P
lluvia	media	normal	--	P
soleado	media	normal	--	P
nublado	media	alta	--	P
nublado	alta	normal	--	P
lluvia	media	alta	--	N

$P(T|sol,alta,N)$

$P(T|sol,nor,P)$

Incertidumbre - RB III, L.E. Suar

18

Nodos ocultos – algoritmo EM

- El algoritmo EM es un método estadístico muy utilizado para estimar probabilidades cuando hay variables no observables (un caso especial es el algoritmo de Baum-Welch en HMM)
- Consiste básicamente de 2 pasos que se repiten en forma iterativa:
 1. Paso E: se estiman los datos faltantes en base a los parámetros (P) actuales
 2. Paso M: se estiman las probabilidades (parámetros) considerando los datos estimados

Incertidumbre - RB III, L.E. Sucar

19

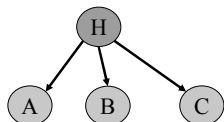
EM para RB con nodos ocultos

1. Iniciar los parámetros desconocidos (CPTs) con valores aleatorios (o estimaciones de expertos)
2. Utilizar los datos conocidos con los parámetros actuales para estimar los valores de la variable(s) oculta(s)
3. Utilizar los valores estimados para completar la tabla de datos
4. Re-estimar los parámetros con los nuevos datos
5. Repetir 2→4 hasta que no haya cambios significativos en las probabilidades

Incertidumbre - RB III, L.E. Sucar

20

Ejemplo



- H es un nodo oculto
- Se seleccionan valores aleatorios para $P(H)$, $P(A|H)$, $P(B|H)$, $P(C|H)$
- Se calcula la probabilidad de H para cada caso, dados los valores de A, B, C
- Cada caso se “pesa” de acuerdo a las probabilidades posteriores de H (un caso puede representar “n” datos)
- Se recalculan los parámetros ($P(H)$, ...) en base a los casos obtenidos
- Se repite el proceso hasta que converja

Incertidumbre - RB III, L.E. Sucar

21

EM

- Limitaciones:
 - Puede caer en máximos locales (depende del valor inicial)
 - Complejidad computacional

Incertidumbre - RB III, L.E. Sucar

22

Aprendizaje Estructural

Diversos métodos:

- Aprendizaje de árboles
- Aprendizaje de poliárboles
- Aprendizaje de redes multiconectadas
 - Métodos basados en medidas
 - Métodos basados en relaciones de dependencia

Incertidumbre - RB III, L.E. Sucar

23

Aprendizaje de árboles

- Algoritmo desarrollado por Chow y Liu para aproximar una distribución de probabilidad por un producto de probabilidades de segundo orden (árbol).
- La probabilidad conjunta de n variables se puede representar como:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{j(i)})$$

- donde $X_{j(i)}$ es la causa o padre de X_i .

Incertidumbre - RB III, L.E. Sucar

24

Aprendizaje de árboles

- Se plantea el problema como uno de optimización - obtener la estructura que más se aproxime a la distribución "real".
- Medida de la diferencia de información entre la distribución real (P) y la aproximada (P*):

$$I(P, P^*) = \sum_x P(X) \log \frac{P(X)}{P^*(X)}$$

- El objetivo es minimizar I.

Incertidumbre - RB III, L.E. Sucar

25

Aprendizaje de árboles

- Se puede definir dicha diferencia en función de la información mutua entre pares de variables, que se define como:

$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i, X_j) \log \frac{P(X_i, X_j)}{P(X_i)P(X_j)}$$

- Se puede demostrar (Chow 68) que la diferencia de información es una función del negativo de la suma de las informaciones mutuas (pesos) de todos los pares de variables que constituyen el árbol
- Encontrar el árbol más próximo equivale a encontrar el árbol con mayor peso.

Incertidumbre - RB III, L.E. Sucar

26

Aprendizaje de árboles - algoritmo

1. Calcular la información mutua entre todos los pares de variables ($n(n-1)/2$).
 2. Ordenar las informaciones mutuas de mayor a menor.
 3. Seleccionar la rama de mayor valor como árbol inicial.
 4. Agregar la siguiente rama mientras no forme un ciclo, si es así, desechar.
 5. Repetir (3-4) hasta que se cubran todas las variables ($n-1$ ramas).
- El algoritmo NO provee la dirección de los arcos, por lo que ésta se puede asignar en forma arbitraria o utilizando semántica externa (experto).

Incertidumbre - RB III, L.E. Sucar

27

Ejemplo (golf)

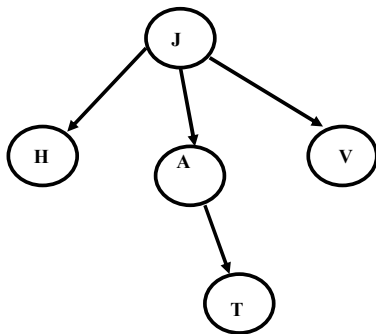
- Informaciones mutuas ordenadas

No.	Var 1	Var 2	I.M.
1	temp.	ambiente	.2856
2	juega	ambiente	.0743
3	juega	humedad	.0456
4	juega	viento	.0074
5	humedad	ambiente	.0060
6	viento	temp.	.0052
7	viento	ambiente	.0017
8	juega	temp.	.0003
9	humedad	temp.	0
10	viento	humedad	0

Incertidumbre - RB III, L.E. Sucar

28

Ejemplo (golf)



Incertidumbre - RB III, L.E. Sucar

29

Aprendizaje de poliárboles

- Parte del esqueleto (estructura sin direcciones) obtenido con el algoritmo anterior
- Determina las dirección de los arcos utilizando pruebas de dependencia entre tripletas de variables.
- Dadas 3 variables, existen 3 casos posibles:
 - Arcos divergentes
 - Arcos secuenciales
 - Arcos convergentes
- Los primeros dos casos son indistinguibles, pero el tercero es diferente, ya que las dos variables "padre" son marginalmente independientes.

Incertidumbre - RB III, L.E. Sucar

30

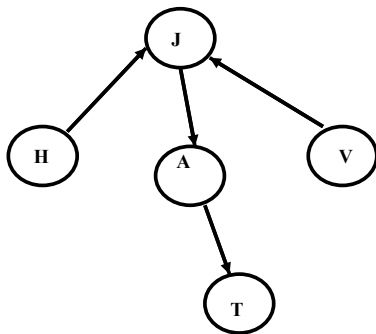
Aprendizaje de poliárboles - algoritmo

1. Obtener esqueleto utilizando el algoritmo de Chow y Liu
2. Recorrer la red hasta encontrar una tripleta de nodos que sean convergentes (tercer caso) - nodo multipadre-
3. A partir de un nodo multipadre determinar las direcciones de los arcos utilizando la prueba de tripletas hasta donde sea posible (base causal).
4. Repetir 2-3 hasta que ya no se puedan descubrir más direcciones.
5. Si quedan arcos sin direccionar utilizar semántica externa para obtener su dirección.

Incertidumbre - RB III, L.E. Sucar

31

Ejemplo



Incertidumbre - RB III, L.E. Sucar

32

Aprendizaje de redes multiconectadas

Existen dos tipos de métodos para el aprendizaje genérico de redes bayesianas:

1. Métodos basados en medidas de ajuste y búsqueda
2. Métodos basados en pruebas de independencia

Incertidumbre - RB III, L.E. Sucar

33

Métodos basados en medidas

Se generan diferentes estructuras y se evalúan respecto a los datos utilizando alguna medida

Dos aspectos principales:

- Medida de “ajuste” de la estructura a los datos
- Búsqueda de la “mejor” estructura

Incertidumbre - RB III, L.E. Sucar

34

Medidas

- Evaluán que tan “buena” es una estructura respecto a los datos
- Hay varias posibles medidas, las dos más comunes son:
 - Medida bayesiana
 - Medida basada en el principio de longitud de descripción mínima (MDL)

Incertidumbre - RB III, L.E. Sucar

35

Medida Bayesiana

- Maximizar la probabilidad de la estructura dados los datos:

$$P(Bs \mid D)$$

- En términos relativos:

$$P(Bs_i \mid D) / P(Bs_j \mid D) = P(Bs_i, D) / P(Bs_j, D)$$

- Considerando variables discretas y que los datos son independientes, las estructuras se pueden comparar en función del número de ocurrencias (frecuencia) de los datos predichos por cada estructura

Incertidumbre - RB III, L.E. Sucar

36

MDL

- La “calidad” de la estructura se basa en el principio de “descripción de longitud mínima” (MDL):
 - Tamaño de la descripción de la red (complejidad)
 - Tamaño de error de predicción de los datos por la red (exactitud)
- Se hace una búsqueda heurística de la estructura en base al MDL

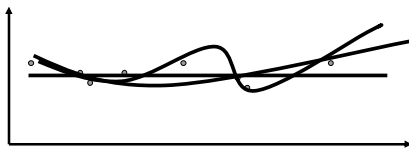
Incertidumbre - RB III, L.E. Sucar

37

MDL

Compromiso entre exactitud y complejidad-
minimizar: *long. de descripción del modelo +*
descripción de lo datos dado el modelo

Ejemplo – ajustar un polinomio a un conjunto de puntos:



Incertidumbre - RB III, L.E. Sucar

38

MDL

Para redes bayesianas:

Complejidad:

$$L = \sum_i [k_i \log_2 n + d(S_i - 1) \prod_{F_i} s_i]$$

n-# de nodos, k-# padres por nodo, Si-# de valores por variable, Fi-conj. de padres, d-# de bits

Exactitud:

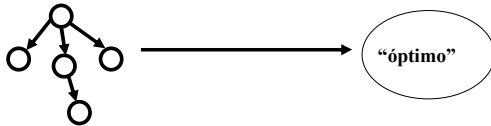
$$w(x_i, Fx_i) = \sum P(x_i, Fx_i) \log_2 [P(x_i, Fx_i)/P(x_i)P(Fx_i)]$$

$$W = \sum_i w(x_i, Fx_i)$$

Incertidumbre - RB III, L.E. Sucar

39

Buscando la mejor estructura

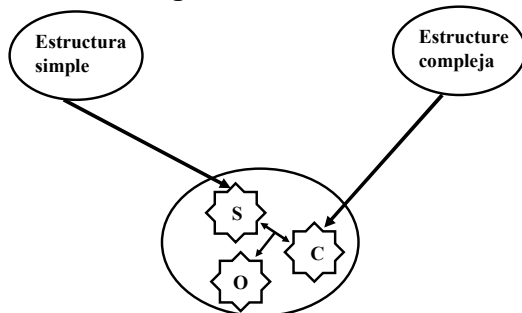


- Búsqueda de ascenso de colinas (*hill climbing*)
- Se inicia con una estructura simple (árbol) y se van agregando arcos hasta llegar a un *mínimo local*

Incertidumbre - RB III, L.E. Sucar

40

Búsqueda bidireccional



Incertidumbre - RB III, L.E. Sucar

41

Algoritmo

- 1 Empezar con la estructura más simple (poliárbol), incrementando la complejidad hasta un mínimo local (*S*)
- 2 Empezar con la estructura compleja (máximos padres), decrementando la complejidad hasta un óptimo local (*C*)
- 3 Obtener la intersección de *S* y *C* y buscar el óptimo global

Incertidumbre - RB III, L.E. Sucar

42

1- Búsqueda “bottom-up”

- Iniciar con una estructura simple (poliárbol)
- Hasta que no haya mejora:
 - Seleccionar la liga con menor MDL
 - Agregar la liga a la estructura actual
 - Verificar por la mejor dirección
 - Eliminar la liga de la lista
- Regresar óptimo “bottom-up” (S)

Incertidumbre - RB III, L.E. Sucar

43

2- Búsqueda “Top-down”

- Iniciar con estructura compleja
- Hasta que no haya mejora:
 - Seleccionar la liga que al quitarla reduce al máximo el MDL
 - Quitar la liga de la estructura actual
- Regresar el óptimo “Top-down” (C)

Incertidumbre - RB III, L.E. Sucar

44

3- Búsqueda de la estructura óptima

- Hacer “óptimo” = intersección de S & C
- Agenda = [óptimo]
- Hasta que no haya mejora en el MDL:
 - Tomar el primero en la agenda y considerar todas las posibles ligas que agregar o remover
 - Ordenar de acuerdo al MDL
 - Tomar las N mejores estructuras
- Regresar la mejor estructura (primera en la agenda): O

Incertidumbre - RB III, L.E. Sucar

45

Parámetros

- Máximo número de padres
- Orden causal (opcional)
- Tamaño del *haz* en la última etapa

Incertidumbre - RB III, L.E. Sucar

46

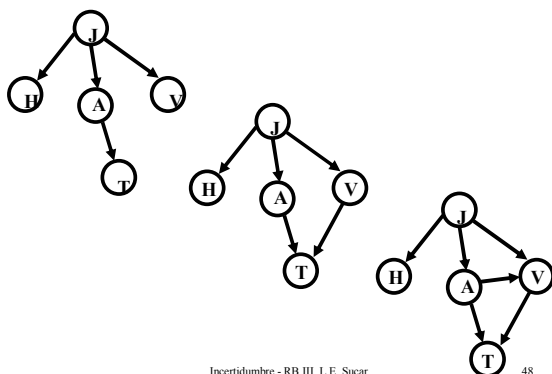
Ejemplo - ¿Cuándo jugar golf?

Ambiente	Temp.	Humedad	Viento	Jugar
soleado	alta	alta	no	N
soleado	alta	alta	si	N
nublado	alta	alta	no	P
lluvia	media	alta	no	P
lluvia	baja	normal	no	P
lluvia	baja	normal	si	N
nublado	baja	normal	si	P
soleado	media	alta	no	N
soleado	baja	normal	no	P
lluvia	media	normal	no	P
soleado	media	normal	si	P
nublado	media	alta	si	P
nublado	alta	normal	no	P
lluvia	media	alta	si	N

Incertidumbre - RB III, L.E. Sucar

47

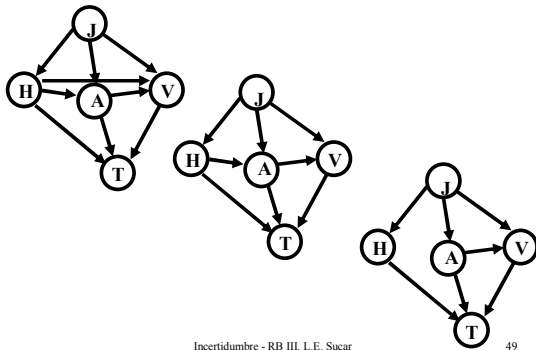
Ejemplo – agregando arcos



Incertidumbre - RB III, L.E. Sucar

48

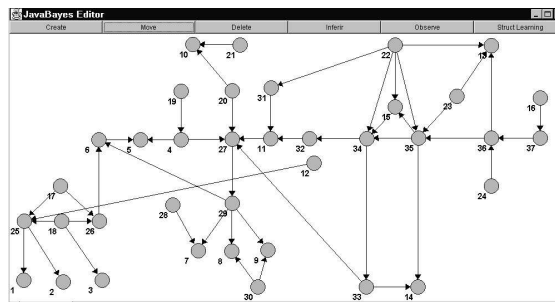
Ejemplo – eliminando arcos



Incertidumbre - RB III, L.E. Sucar

49

Ejemplo - Red ALARM



Incertidumbre - RB III, L.E. Sucar

50

Métodos basados en medidas

- Se genera la estructura en base a ir agregando arcos de acuerdo a medidas de dependencia entre variables
- Ejemplos:
 - Árboles – método de Chow y Liu
 - Poliárboles – método de Rebane y Pearl
 - Multiconectadas – existen varios algoritmos basados en diferentes medidas y que dependen del ordenamiento de variables (orden causal)

Incertidumbre - RB III, L.E. Sucar

51

Combinación de conocimiento y datos

- Restricciones:
 - Se incorpora conocimiento previo a los algoritmos de aprendizaje estructural
 - Por ejemplo:
 - Orden de las variables (orden causal)
 - Dependencias conocidas
 - Independencias conocidas

Incertidumbre - RB III, L.E. Sucar

52

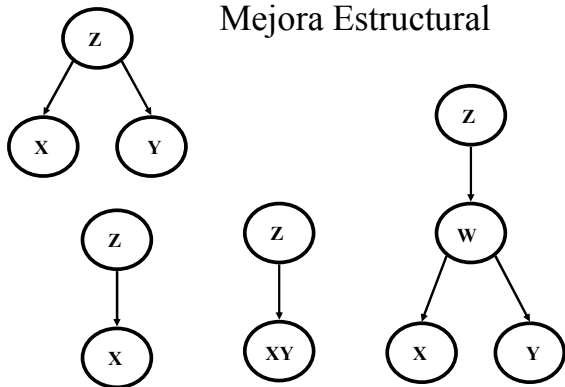
Combinación de conocimiento y datos

- Mejora:
 - Se parte de una estructura dada por un experto (subjetiva) y se mejora con datos
 - Por ejemplo, verificando relaciones de independencia y alterando la estructura:
 - Eliminar nodos
 - Combinar nodos
 - Insertar nodos

Incertidumbre - RB III, L.E. Sucar

53

Mejora Estructural



Incertidumbre - RB III, L.E. Sucar

54

Referencias

- Pearl 88 – Cap. 8
- Neapolitan 90 – Cap. 10
- T. Mitchell, Machine Learning, McGraw-Hill, 1997 – Cap. 6
- Borglet & Kruse, Graphical Models, Wiley – Cap. 5 (EM)

Incertidumbre - RB III, L.E. Sucar

55

Referencias

- W. Lam, F. Bacchus, "Learning Bayesian Belief Networks: An Approach based on the MDL Principle", Computational Intelligence, Vol. 10 (1994) 269-293.
- G. Cooper, E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data", Machine Learning, Vol 9, 1992.
- L. E. Sucar, D. F. Gillies, D. A. Gillies, "Objective Probabilities in Expert Systems", Artificial Intelligence Journal, Vol. 61 (1993) 187-208.
- W. Buntine, "A guide to the literature on learning probabilistic networks from data", IEEE TKDE.

Incertidumbre - RB III, L.E. Sucar

56

Actividades

- Hacer ejercicios de aprendizaje de redes bayesianas

Incertidumbre - RB III, L.E. Sucar

57
