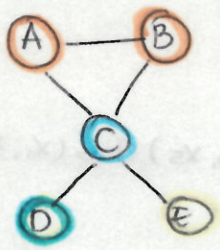
Estadística Multivariada & Datos Categóricos 04/may/14

Tarea #5

Alejandra Celo de Larrea Ibarra 124433

1. Definan un grafo matemático con vértices $\{A, B, C, D, E\}$ tal que los conjuntos $\{A, B\}$ & $\{E\}$ sean separados por el conjunto $\{C, D\}$.



Nodos = $\{A, B, C, D, E\}$

Bordes = $\{\{A,B\}, \{A,C\}, \{B,C\}, \{C,D\}, \{C,E\}\}$

2. Definan un grafo matemático con vértices $\{A, B, C, D, E\}$ tal que los conjuntos $\{A, B\}$ & $\{E\}$ sean separados por el conjunto $\{C\}$ pero no por el conjunto $\{C, D\}$

No hay solución. Dado que $\{C\} \subset \{C, D\} \subset V$, si $\{C\}$ separa al conjunto $\{A, B\}$ del conjunto $\{E\}$, entonces $\{C, D\}$ también los separa. Se necesitaría un sexto nodo F para poder conectar los conjuntos $\{A, B\}$ & $\{E\}$ "sin pasar" por $\{C, D\}$. & $\{C\}$.

3. Escriban los elementos del grafo probabilístico saturado de una tabla de contingencias de 5 dimensiones, donde c/ dimensión tiene 2 categorías (ie modelo log-lineal saturado 2^5)

variables: x_1, x_2, x_3, x_4, x_5 con $x_i \in \mathcal{X}_i = (e_{x_i}(1), e_{x_i}(2))$

vértices: $\{1, 2, 3, 4, 5\}$

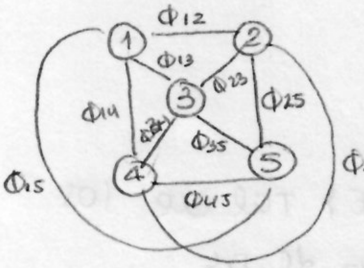
bordes: $\{\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{4,5\}\}$

opción 1:

$$P(x_1, x_2, x_3, x_4, x_5) = \Phi_{12}^{(x_1,x_2)} \Phi_{13}^{(x_1,x_3)} \Phi_{14}^{(x_1,x_4)} \Phi_{15}^{(x_1,x_5)} \Phi_{23}(x_2,x_3)\, \Phi_{2,4}(x_2,x_4)\, \Phi_{2,5}(x_2,x_5)\, \Phi_{3,4}(x_3,x_4)$$
$$\Phi_{3,5}(x_3,x_5)\, \Phi_{4,5}(x_4,x_5)$$

cuyo grafo correspondiente es:

y con modelo loglineal dado por



$$\log(\theta_v) = \log(P(x_v)) = \log(P(x_1 = e(1), x_2 = e(2), x_3 = e(3), x_4 = e(4), x_5 = e(5)))$$

$$= a^0 + a^{1,2}_{e(1,2)} + a^{1,3}_{e(1,3)} + a^{14}_{e(1,4)} + a^{1,5}_{e(1,5)} + a^{2,3}_{e(2,3)}$$

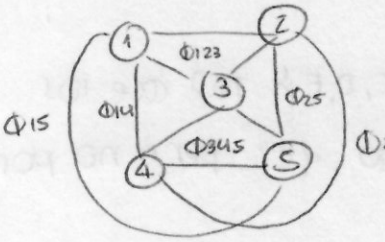
$$+ a^{2,4}_{e(2,4)} + a^{2,5}_{e(2,5)} + a^{3,4}_{e(3,4)} + a^{3,5}_{e(3,5)} + a^{4,5}_{e(4,5)}$$

opción 2:

$$P(x_1, x_2, x_3, x_4, x_5) = \Phi_{1,2,3}(x_1, x_2, x_3)\,\Phi_{1,4}(x_1, x_4)\,\Phi_{2,5}(x_2, x_5)\,\Phi_{3,4,5}(x_3, x_4, x_5)\,\Phi_{15}(x_1, x_5)\,\Phi_{2,4}(x_1, x_4)$$

cuyo grafo correspondiente es

y con modelo loglineal dado por



$$\log(\theta_v) = \log(P(x_v)) = \log(P(x_1 = e(1), x_2 = e(2), x_3 = e(3), x_4 = e(4), x_5 = e(5)))$$

$$= a^0 + a^{14}_{e(1,4)} + a^{15}_{e(1,5)} + a^{2,5}_{e(2,5)} + a^{2,4}_{e(2,4)} + a^{1,2,3}_{e(1,2,3)} + a^{3,4,5}_{e(3,4,5)}$$

4. Escriban los elementos del grafo probabilístico independiente de una tabla de contingencias de 5 dimensiones donde c/ dimensión tiene 2 categorías.

variables := $\{x_1, x_2, x_3, x_4, x_5\}$ con $x_i \in \mathcal{X}_i = \{e_{x_i}(1), e_{x_i}(2)\}$

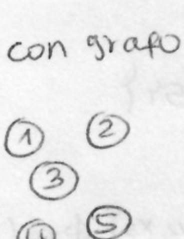
vértices $= \{1, 2, 3, 4, 5\}$

Bordes $= \emptyset$

$$\theta_v = P(x_1, x_2, x_3, x_4, x_5) = \Phi_1(x_1)\,\Phi_2(x_2)\,\Phi_3(x_3)\,\Phi_4(x_4)$$

con grafo asociado

y con modelo log-lineal



$$\log(\theta_v) = a^1_{e(1)} + a^2_{e(2)} + a^3_{e(3)} + a^4_{e(4)} + a^5_{e(5)}$$

$$= \log(P(x_1 = e(1), x_2 = e(2), x_3 = e(3), x_4 = e(4), x_5 = e(5)))$$

5. Respecto a los modelos loglineales, contrasten para el ejemplo de los reptiles, (tabla de contingencia) las probabilidades muestrales del modelo independiente contra las probabilidades muestrales. ie identificar si las probabilidades muestrales pueden expresarse como el producto de las probabilidades marginales de las tres dimensiones.

La tabla de contingencias es:

X_3 = "<=4.75"

X_1 \ X_2	"<=4"	">4"
anoli	86	35
dist	73	70

X_3 = ">4.75"

X_1 \ X_2	"<=4"	">4"
anoli	32	11
dist	61	41

$\Rightarrow$

marginales

$p(X_1=\text{"anoli"}) = \frac{86+35+32+11}{409} = \frac{164}{409} = 0.4009780$

$p(X_1=\text{"dist"}) = \frac{73+70+61+41}{409} = \frac{245}{409} = 0.5990220$

$p(X_2=\text{"<=4"}) = \frac{86+73+32+61}{409} = \frac{252}{409} = 0.6161369$

$p(X_2=\text{">4"}) = \frac{35+70+11+41}{409} = \frac{157}{409} = 0.3838631$

$p(X_3=\text{"<=4.75"}) = \frac{86+35+73+70}{409} = \frac{264}{409} = 0.6454768$

$p(X_3=\text{">4.75"}) = \frac{32+11+61+41}{409} = \frac{145}{409} = 0.3545232$

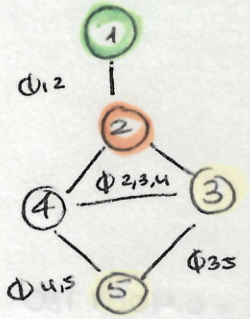
Comparaciones:

X_1	X_2	X_3	$P(X_1, X_2, X_3)$	$P(X_1=x_1)P(X_2=x_2)P(X_3=x_3)$
anoli	<=4	≤4.75	$\frac{86}{409}$ = 0.2102	(0.4009780)(0.6161369)(0.6454768) = 0.1594
anoli	<=4	>4.75	$\frac{32}{409}$ = 0.0782	(0.4009780)(0.6161369)(0.3545232) = 0.0875
anoli	>4	≤4.75	$\frac{35}{409}$ = 0.0855	(0.4009780)(0.3838631)(0.6454768) = 0.0993
anoli	>4	>4.75	$\frac{11}{409}$ = 0.0268	(0.4009780)(0.3838631)(0.3545232) = 0.0545
dist	<=4	≤4.75	$\frac{73}{409}$ = 0.1784	(0.5990220)(0.6161369)(0.6454768) = 0.2382
dist	<=4	>4.75	$\frac{61}{409}$ = 0.1491	(0.5990220)(0.6161369)(0.3545232) = 0.13084
dist.	>4	≤4.75	$\frac{70}{409}$ = 0.1711	(0.5990220)(0.3838631)(0.6454768) = 0.1484
dist.	>4	>4.75	$\frac{41}{409}$ = 0.1002	(0.5990220)(0.3838631)(0.3545232) = 0.0815

$\therefore$ El modelo no tiene evidencia de independencia.

6. Para el grafo g3 del ejemplo visto en clase, identificaren si $\{x_1\}$ es condicionalmente independiente de $\{x_3, x_5\}$ dado $\{x_2\}$

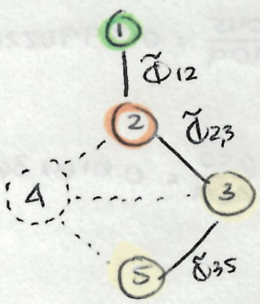
Si $\Phi_V = \Phi_{A,S} \times \Phi_{B,S}$ entonces $(A \perp B) | S$



sea $\Phi_V = \Phi_{12}(x_1, x_2)\, \Phi_{2,3,4}(x_2, x_3, x_4)\, \Phi_{4,5}(x_4, x_5)\, \Phi_{3,5}(x_3, x_5)$

marginalizando sobre x_4, tenemos

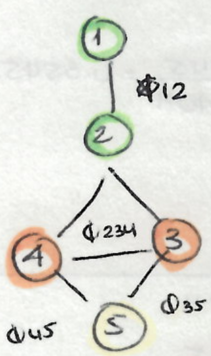
$$\Phi_V = \underbrace{\tilde{\Phi}_{12}(x_1, x_2)}_{\Phi_{A,S}}\, \underbrace{\tilde{\Phi}_{2,3}(x_2, x_3)\, \tilde{\Phi}_{3,5}(x_3, x_5)}_{\Phi_{B,S}}$$



$\therefore (\{x_1\} \perp \{x_3, x_5\}) | \{x_2\}$ cuando marginalizamos sobre x_4.

7. Para el mismo grafo g3, identificaren si $\{x_1, x_2\}$ es condicionalmente independiente de $\{x_5\}$ dado $\{x_3, x_4\}$.

Si $\Phi_V = \Phi_{A,S} \times \Phi_{B,S}$ entonces $(A \perp B) | S$



sea

$$\Phi_V = \underbrace{\Phi_{12}(x_1, x_2)\, \Phi_{234}(x_2, x_3, x_4)}_{\Phi_{A,S}}\, \underbrace{\Phi_{4,5}(x_4, x_5)\, \Phi_{3,5}(x_3, x_5)}_{\Phi_{B,S}}$$

$\therefore (\{x_1, x_2\} \perp \{x_5\}) | \{x_4, x_5\}$ //