Modelos Gráficos Probabilistas

L. Enrique Sucar
INAOE

Sesión 5:
Métodos Básicos

“ ... tenemos razones para creer que hay en la constutución de las cosas leyes de acuerdo a las cuales suceden los eventos ...”

[Richard Price, 1763]

Métodos Básicos

- Formulación
- Probabilidad conjunta
- Cálculo directo (*fuerza bruta*):
  - Probabilidades marginales / condicionales
  - Eventos más probables
  - Estimación de probabilidades
- Análisis

Formulación

- Muchos problemas se pueden formular como un conjunto de variables sobre las que tenemos cierta información y queremos obtener otra, por ejemplo:
  - Diagnóstico médico o industrial (síntomas, enfermedades, fallas, …)
  - Percepción (sensores, imágenes, señales, objetos, …)
  - Clasificación (datos bancarios, datos estudiantes, …)
  - Modelado de usuarios (interacciones, emociones, …)

Ejemplo

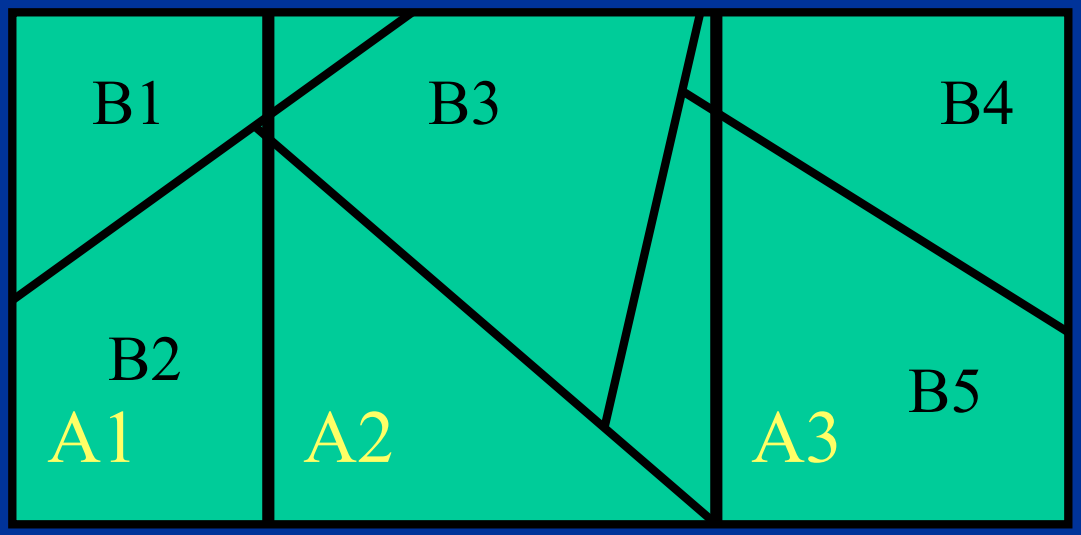
- Determinar si una persona es sujeta de crédito:
  - X1: otorgar crédito (si/no)
  - X2: ingreso anual (entero positivo)
  - X3: créditos anteriores (si/no)
  - X4: edad (entero positivo)
  - X5: ocupación (empleado, empresario, …)

Formulación

- Desde el punto de vista de probabilidad se puede ver como:
  - Un conjunto de variables aleatorias: X1, X2, X3, ...
  - Cada variable es generalmente una partición del espacio
  - Cada variable tiene una distribución de probabilidad (conocida o desconocida)

Variables y Particiones

- A = {A1, A2, A3}
- B = {B1, B2, B3, B4, B5}



Preguntas

- Dada cierta información (como valores de variables y probabilidades), se requiere contestar ciertas preguntas, como:
  - Probabilidad de que una variable tome cierto valor [marginal *a priori*]
  - Probabilidad de que una variable tome cierto valor dada información de otra(s) variable(s) [condicional o *a posteriori*]

Preguntas

- Valor de mayor probabilidad de una o más variables [abducción]
- Valor de mayor probabilidad de una o más variables dada información de otra(s) variable(s) [abducción parcial o explicación]
- Parámetros del modelo dados datos históricos de las variables [estimación o aprendizaje]

Enfoque básico (*fuerza bruta*)

- Dada la probabilidad conjunta de las variables, para todos los posibles valores de cada una (asumimos por ahora que son discretas):

$$P(X1, X2, X3, ..., Xn)$$

- podemos estimar todas las probabilidades requeridas

Inferencia

- Probabilidad marginal (cuál es la probabilidad de las diferentes ocupaciones):

$$p(X) = \Sigma_{Y, Z}\, p(X,Y, Z)$$

- Probabilidad condicional (cuál es la probabilidad de otorgar el crédito dado cierto nivel de ingreso) :

$$p(X \mid Y) = p(X,Y) / p(Y)$$

- Donde:

$$p(X,Y) = \Sigma_{Z}\, p(X,Y, Z)$$

Abducción

- Valor más probable (qué tipo de ocupación es el más común):

$$\text{Arg}_X \, [\max p(X) = \max \Sigma_{Y,Z} \, p(X,Y,Z)]$$

- Valor condicional más probable (debo o no otorgar el crédito):

$$\text{Arg}_X \, [\max p(X \mid y1) = \max p(X,y1) / p(y1)]$$

- Valor conjunto más probable (que combinación de ocupación y edad es la más probable):

$$\text{Arg}_{X,Y} \, [\max p(X,Y) = \max \Sigma_{Z} \, p(X,Y,Z)]$$

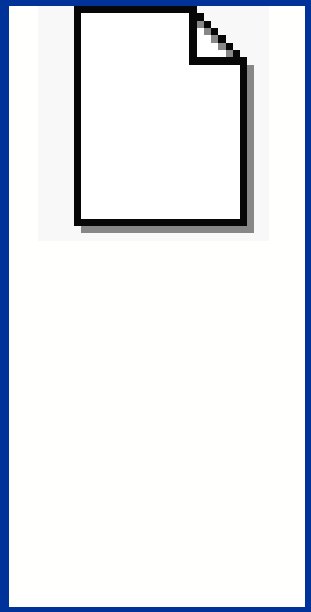
Ejemplo

- Problema de decidir cuando jugar golf?
- Variables
  - Ambiente
  - Temperatura
  - Viento
  - Humedad
  - Jugar

Ejemplo

- Consideremos inicialmente dos variables: ambiente (S,N,Ll) y temperatura (A,M,B)
- Dada la tabla de P conjunta, encontrar:
  - Probabilidad de ambiente, temperatura
  - Probabilidad de ambiente conocida la temperatura (y viceversa)
  - Combinación de A y T más probable
  - Ambiente más probable dada la temperatura (y viceversa)

Ejemplo



Limitaciones

- El tamaño de la tabla y el número de operaciones crece exponencialmente con el número de variables (complejidad computacional)
- La "tabla" conjunta nos dice poco sobre el fenómeno que estamos analizando (limitaciones cognitivas)
- Es difícil estimar las probabilidades requeridas, ya sea por expertos o a partir de datos (complejidad estadística)

Estimación de Parámetros

- Dados un conjunto de valores de las variables (registros), se busca estimar las probabilidades conjuntas requeridas
- Considerando datos completos:
  - Las probabilidades se pueden *estimar* contando el número de casos de cada valor

$$P(Xi,Yj) \sim Ni,j / N$$

  - Esto corresponde al *estimador de máxima verosimilitud* cuando no hay valores faltantes

Ejemplo

- Dados datos sobre lo que “jugadores” han hecho en situaciones pasadas, podemos estimar la probabilidad conjunta
- Consideremos el caso de 2 variables (ambiente y temperatura) y 14 registros de datos

Ejemplos

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

Ejemplo

Limitaciones

- Se requiere una gran cantidad de datos para estimaciones confiables
- Se complica si hay datos faltantes
- Puede ser mejor estimar probabilidades marginales o condicionales (menos datos, más fácil para el experto)

Análisis

- El problema de complejidad computacional utilizando el enfoque básico (tanto en espacio para representar el modelo, como en tiempo para el cálculo de probabilidades), nos lleva a buscar alternativas
- Los modelos gráficos probabilistas proveen esta alternativa, mediante representaciones mucho más compactas (y entendibles) y técnicas eficientes para el cálculo de las probabilidades

Referencias

- [Koller y Friedman] Cap. 2

Actividades

- Ejercicios métodos básicos