

Sesión 4: Métodos Probabilísticos Básicos

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

[Richard Price, 1763]

Métodos Básicos

- Probabilidad conjunta
- Cálculo directo (*fuerza bruta*):
 - Probabilidades marginales / condicionales
 - Eventos más probables
- Estimación directa
- Clasificación
 - Clasificador bayesiano simple
 - Otros clasificadores
 - Regresión

Incertidumbre - Métodos Básicos -
L.E. Sucar

2

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
 - Percepción (visión, voz, sensores)
 - Clasificación (bancos, empleadores, ...)
 - Modelado de estudiantes, usuarios, etc.

Incertidumbre - Métodos Básicos -
L.E. Sucar

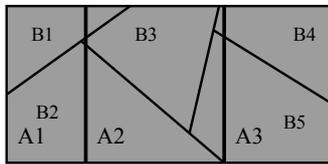
3

Formulación

- Desde el punto de vista de probabilidad se puede ver como:
 - Un conjunto de variables aleatorias: X_1, X_2, X_3, \dots
 - Cada variable es generalmente una partición del espacio
 - Cada variable tiene una distribución de probabilidad (conocida o desconocida)

Variables y Particiones

- $A = \{A_1, A_2, A_3\}$
- $B = \{B_1, B_2, B_3, B_4, B_5\}$



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]
- Dados datos históricos de las variables estimar sus probabilidades [estimación o aprendizaje]

Enfoque básico (*fuerza bruta*)

- Dada la probabilidad conjunta de las variables podemos estimar todas las probabilidades requeridas

$$P(X_1, X_2, X_3, \dots, X_n)$$

- Para todos los posibles valores de cada variable (asumimos por ahora que son discretas)

Inferencia

- Probabilidad marginal:
 $p(X) = \sum_{Y, Z} p(X, Y, Z)$
- Probabilidad condicional:
 $p(X | Y) = p(X, Y) / p(Y)$
- Donde:
 $p(X, Y) = \sum_Z p(X, Y, Z)$

Abducción

- Valor más probable:
 $\text{Arg}_X [\max p(X) = \max \sum_{Y, Z} p(X, Y, Z)]$
- Valor condicional más probable:
 $\text{Arg}_X [\max p(X | y1) = \max p(X, y1) / p(y1)]$
- Valor conjunto más probable:
 $\text{Arg}_{X, Y} [\max p(X, Y) = \max \sum_Z p(X, Y, Z)]$

Incertidumbre - Métodos Básicos -
L.E. Sucar

10

Ejemplo

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

Incertidumbre - Métodos Básicos -
L.E. Sucar

11

Ejemplo

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

Incertidumbre - Métodos Básicos -
L.E. Sucar

12

Ejemplo



Hoja de cálculo de
Microsoft Excel

Limitaciones

- El tamaño de la tabla y el número de operaciones crece exponencialmente con el número de variables
- La “tabla” conjunta nos dice poco sobre el fenómeno que estamos analizando
- Puede ser difícil estimar las probabilidades requeridas (por expertos o a partir de datos)

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
$$P(X_i, Y_j) \sim N_{i,j} / N$$
 - Esto corresponde al *estimador de máxima verosimilitud* cuando que no hay valores faltantes

Ejemplo

- 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



Hoja de cálculo de
Microsoft Excel

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)
- También puede ser complejo el tener *demasiados* datos (minería de datos)

Clasificación

- El concepto de clasificación tiene dos significados:
 - No supervisada: dado un conjunto de datos, establecer clases o agrupaciones (clusters)
 - Supervisada: dadas ciertas clases, encontrar una regla para clasificar una nueva observación dentro de las clases existentes

Clasificación

- El problema de clasificación (supervisada) consiste en obtener el valor más probable de una variable (hipótesis) dados los valores de otras variables (evidencia, atributos)

$$\text{Arg}_H [\text{Max } P(H | E_1, E_2, \dots, E_N)]$$
$$\text{Arg}_H [\text{Max } P(H | \mathbf{E})]$$

$$\mathbf{E} = \{E_1, E_2, \dots, E_N\}$$

Tipos de Clasificadores

- Métodos estadísticos clásicos
 - Clasificador bayesiano simple (*naive Bayes*)
 - Discriminadores lineales
- Modelos de dependencias
 - Redes bayesianas
- Aprendizaje simbólico
 - Árboles de decisión, reglas
- Redes neuronales

Incertidumbre - Métodos Básicos -
L.E. Sucar

22

Clasificación

- Consideraciones para un clasificador:
 - Exactitud – proporción de clasificaciones correctas
 - Rapidez – tiempo que toma hacer la clasificación
 - Claridad – que tan comprensible es para los humanos
 - Tiempo de aprendizaje – tiempo para obtener o ajustar el clasificador a partir de datos

Incertidumbre - Métodos Básicos -
L.E. Sucar

23

Regla de Bayes

- Para estimar esta probabilidad se puede hacer en base a la regla de Bayes:

$$P(H | E) = P(H) P(E | H) / P(E)$$

$$P(H | E) = P(H) P(E | H) / \sum_i P(E | H_i) P(H_i)$$

- Normalmente no se requiere saber el valor de probabilidad, solamente el valor más probable de H

Incertidumbre - Métodos Básicos -
L.E. Sucar

24

Regla de Bayes

- Para el caso de 2 clases $H: \{0, 1\}$, la regla de decisión de Bayes es:

$$H^*(E) = \begin{cases} 1 & \text{si } P(H=1 | E) > 1/2 \\ 0, & \text{de otra forma} \end{cases}$$

- Se puede demostrar que la regla de Bayes es óptima

Valores Equivalentes

- Se puede utilizar cualquier función monotónica para la clasificación:

$$\text{Arg}_H [\text{Max } P(H | E)]$$

$$\text{Arg}_H [\text{Max } P(H) P(E | H) / P(E)]$$

$$\text{Arg}_H [\text{Max } P(H) P(E | H)]$$

$$\text{Arg}_H [\text{Max } \log \{P(H) P(E | H)\}]$$

$$\text{Arg}_H [\text{Max } \log P(H) + \log P(E | H)]$$

Clasificador bayesiano simple

- Estimar la probabilidad: $P(E | H)$ es complejo, pero se simplifica si se considera que los atributos son independientes dada la hipótesis:

$$P(E_1, E_2, \dots, E_N | H) = P(E_1 | H) P(E_2 | H) \dots P(E_N | H)$$

- Por lo que la probabilidad de la hipótesis dada la evidencia puede estimarse como:

$$P(H | E_1, E_2, \dots, E_N) = \frac{P(H) P(E_1 | H) P(E_2 | H) \dots P(E_N | H)}{P(E)}$$

- Esto se conoce como el clasificador bayesiano simple

Clasificador bayesiano simple

- Como veíamos, no es necesario calcular el denominador:

$$P(H | E_1, E_2, \dots, E_N) \sim P(H) P(E_1 | H) P(E_2 | H) \dots P(E_N | H)$$

- $P(H)$ se conoce como la probabilidad a priori, $P(E_i | H)$ es la probabilidad de los atributos dada la hipótesis (verosimilitud), y $P(H | E_1, E_2, \dots, E_N)$ es la probabilidad posterior

Ejemplo

- Para el caso del golf, cuál es la acción más probable (jugar / no-jugar) dado el ambiente y la temperatura?



Hoja de cálculo de
Microsoft Excel

Ventajas

- Bajo tiempo de clasificación
- Bajo tiempo de aprendizaje
- Bajos requerimientos de memoria
- “Sencillez”
- Buenos resultados en muchos dominios

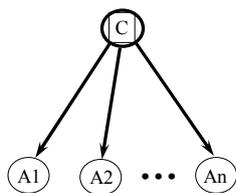
Limitaciones

- En muchas ocasiones la suposición de independencia condicional no es válida
- Para variables continuas, existe el problema de discretización
- Alternativas:
 - Probabilidad conjunta (complejidad)
 - Discriminador lineal (variables gaussianas)
 - Considerar algunas dependencias (redes bayesianas)

Incertidumbre - Métodos Básicos -
L.E. Sucar

31

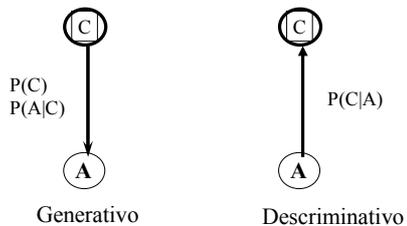
CBS – modelo gráfico



Incertidumbre - Métodos Básicos -
L.E. Sucar

32

Enfoques para clasificación

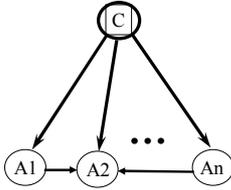


Incertidumbre - Métodos Básicos -
L.E. Sucar

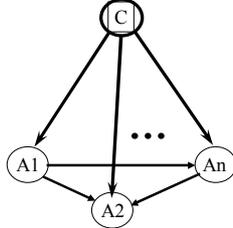
33

Extensiones

• TAN



• BAN



Incertidumbre - Métodos Básicos -
L.E. Sucar

34

Discriminador lineal

- Se define un hiperplano (discriminante) que es una combinación lineal de los atributos:

$$g(X) = \sum a_j x_j,$$

x_j - promedios de clase,

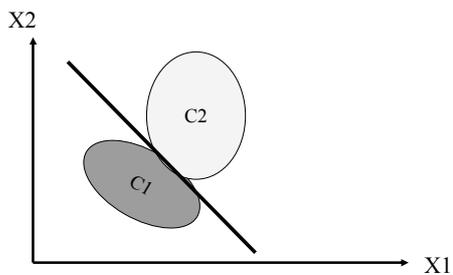
$a_1 \dots a_n$ - coeficientes

- Asumiendo una distribución normal multivariada, se puede obtener la ecuación del hiperplano en función de los promedios y covarianzas de las clases

Incertidumbre - Métodos Básicos -
L.E. Sucar

35

Discriminador lineal



Incertidumbre - Métodos Básicos -
L.E. Sucar

36

Descriminador Lineal

- Para el caso gaussiano, la probabilidad posterior es una función logística (rampa):
$$P(C_n | A_n) = 1 / [1 + \exp(-\theta^T A_n)]$$
- Donde el parámetro θ depende de las medias y covarianzas de las distribuciones condicionales de cada clase
- Ejemplo en 1-D

Costo de mala clasificación

- En realidad, no sólo debemos considerar la clase más probable si no también el costo de una mala clasificación
 - Si el costo es igual para todas las clases, entonces es equivalente a seleccionar la de mayor probabilidad
 - Si el costo es diferente, entonces se debe minimizar el costo esperado

Referencias

- D. Michie, D.J. Spiegelhalter, C.C. Taylor, "Machine Learning, Neural and Statistical Classification", Ellis Horwood, 1994
- [Notas Jordan] – Capítulo 5
- J. Cheng, R. Greiner, "Comparing Bayesian network classifiers", UAI'99, 101-108.
- Libros básicos de probabilidad, por ej.:
 - Meyer, Introductory Probability and Statistical Applications
 - Wasserman, All of Statistics, Springer

Actividades

- Implementar clasificador bayesiano simple en MatLab (estimación de parámetros y cálculo de probabilidades)
- Probar con datos de Golf
- Probar con otras bases de datos
