

Sistemas Inteligentes

Dpnt. de Ciència de la Computació i Intel·ligència Artificial Dpto. de Ciencia de la Computación e Inteligencia Artificial

Aprendizaje bayesiano

- Basado en el **teorema de Bayes**
- •Permite combinar los datos de ejemplo con conocimiento a priori
- •Usado como clasificador, puede obtener probabilidades de pertenecer a cada clase



•Posibilidad de construir representaciones más complejas (Modelos de ocultos de Markov, redes bayesianas, etc).

Tema 8. Aprendizaje Bayesiano

Dpnt. de Ciência de la Computació i Intel·ligència *A*rtifici Opto. de Ciencia de la Computación e Inteligencia *A*rtific

Fundamentos. Teorema de Bayes

•Aprendizaje bayesiano: buscar la hipótesis *h* (de entre todas las H posibles) más probable si hemos observado una serie de datos *D* (*máximo a posteriori o MAP*)

$$h_{MAP} \equiv argmaxP(h\mid D)$$

Base: Teorema de Bayes

Universitat d'Alacant
Universidad de Alicante

 $P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$

Tema 8. Aprendizaje Bayesiano

3



Fundamentos. MAP y ML

 $h_{MAP} \equiv argmaxP(h \mid D)$

Máximo a posteriori o MAP

$$h_{MAP} = argmax_{h \in H} \frac{P(D \mid h)P(h)}{P(D)}$$

 $h_{MAP} = argmax_{h \in H} P(D \mid h) P(h)$

ya que P(D) cte. independiente de h

Si además suponemos P(h)=cte

(a priori, todas las hipótesis son igualmente probables), entonces

 $h_{ML} = argmax_{h \in H} P(D \mid h)$

Máxima verosimilitud o

maximum likelihood (ML)

Tema 8. Aprendizaje Bayesiano

. de Ciència de la Computació i Intel·ligència *d*rtifició . de Ciencia de la Computación e Inteligencia *d*rtifició

Clasificador bayesiano

- Las hipótesis son las clases a las que puede pertenecer un ejemplo
- •Suponemos ejemplos caracterizados como tuplas de atributos $\langle a_1, a_2, \dots a_n \rangle$

$$C_{MAP} \equiv \underset{c_i \in C}{argmaxP}(c_i \mid a_{1,}a_{2}...a_n)$$

$$C_{MAP} \equiv \underset{c_i \in C}{argmax} \frac{P \left(a_{1,} a_{2} \dots a_{n} \mid c_i \right) P \left(c_i \right)}{P \left(a_{1,} a_{2} \dots a_{n} \right)}$$

$$C_{MAP} \equiv \underset{c_i \in C}{argmax} P(a_1, a_2 \dots a_n \mid c_i) P(c_i)$$

Universitat d'Al
Universidad de

Tema 8. Aprendizaje Bayesiano

5

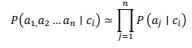
Sistemas Inteligentes

Estimar las probabilidades

Dpnt. de Ciència de la Computació i Intel·ligència *A*rt Dpto. de Ciencia de la Computación e Inteligencia *A*r

 $C_{MAP} \equiv \underset{c_{i} \in C}{argmaxP(a_{1,}a_{2} \dots a_{n} \mid c_{i})P(c_{i})} \qquad \underset{n^{0} \text{ deejemplos delaclase i}}{\underbrace{n^{0} \text{ deejemplos conatributos } a_{1,}a_{2} \dots a_{n}}} \qquad \underset{n^{0} \text{ demasiado pequeño}}{\underbrace{n^{0} \text{ demasiado pequeño}}}$

Simplificación: suponemos que los valores de los atributos son condicionalmente independientes para una clase dada (naive bayes classifier)



Universitat d'Alacant Universidad de Alicant

Tema 8. Aprendizaje Bayesiano

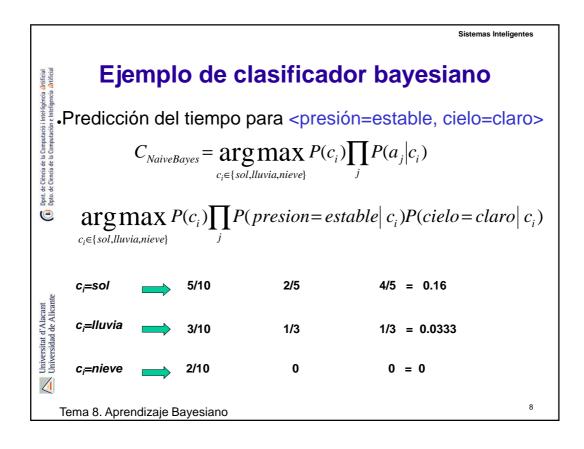
Sistemas Inteligentes

Universitat d'Alacant Universidad de Alicante

Ejemplo: predicción del tiempo

DÍA	TEMP.	DIR. VIENTO	CIELO	Presión	Tiempo
1	≤ 0	Sur	Nuboso	Subiendo	Sol
2	> 0	Oeste	Claro	Estable	Sol
3	> 0	Norte	Claro	Subiendo	Sol
4	> 0	Norte	Claro	Bajando	Lluvia
5	> 0	Oeste	Nuboso	Bajando	Lluvia
6	≤ 0	Norte	Nuboso	Bajando	Nieve
7	> 0	Sur	Nuboso	Estable	Lluvia
8	> 0	Sur	Claro	Subiendo	Sol
9	≤ 0	Este	Nuboso	Bajando	Nieve
10	≤ 0	Sur	Claro	Estable	Sol

Tema 8. Aprendizaje Bayesiano



Dpnt. de Ciència de la Computació i Intel·ligència dat Dpto. de Ciencia de la Computación e Inteligencia da

Ejemplo 2: clasificador de textos

- •Clasificar un texto en una categoría predefinida, dados:
- El conjunto de palabras del texto (atributos)

En un lugar de la Mancha, de cuyo nombre no quiero acordarme, no ha mucho tiempo que vivía un hidalgo de los de lanza en astillero, adarga antigua, rocín flaco y galgo corredor

 $a_1=en$, $a_2=un$... $a_{32}=corredor$

Universitat d'Alacant Universidad de Alicar

.El conjunto de posibles categorías C

c₁=interesante, c₂=no-interesante

Tema 8. Aprendizaje Bayesiano

9

Sistemas Inteligentes

Dpnt. de Ciència de la Computació i Intel·ligència Art Dpto. de Ciencia de la Computación e Inteligencia Ar

Simplificaciones

•Clasificador bayesiano "naive": independencia de los atributos

$$P(a_1, a_2 ... a_n | c_i) = \prod_{j=1}^n P(a_j | c_i)$$

 No importa la posición de las palabras en el texto, solo si están presentes

 $P(a_{31} = galgo \mid interesante) \simeq P(galgo \mid interesante)$

Sustituimos $P(a_i | c_i)$ por $P(w_i | c_i)$

Universitat d'Alacan Universidad de Alica

Tema 8. Aprendizaje Bayesiano



Universitat d'Alacant Universidad de Alicante

Fase 1: Aprendizaje

- -Tomar un conjunto de ejemplos $x_i \in X$ etiquetados con las clases a las que pertenecen
- Voc = conjunto de palabras en X (sin considerar preposiciones, artículos, etc.)
- Para cada clase c_i calcular $P(c_i)$, $P(w_k | c_i)$ como sigue:
- Calcular $docs_i$, documentos de la clase c_i
 - Calcular prob. a priori de la clase c_j $P(c_j) = \frac{|docs_j|}{|X|}$
- Calcular

-text_i, concatenación de todos los docs. de docs_i

- -n, número de posiciones en text_i
- $-n_k$, número de veces que aparece w_k en $text_j$
- •Calcular $P(w_k \mid c_j) = \frac{n_k + 1}{n + |Voc|}$

Tema 8. Aprendizaje Bayesiano

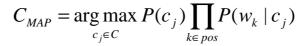
11



Sistemas Inteligentes

Fase 2 - Clasificación

- •Dado un documento $x=w_1, w_2, ..., w_n$
- •Quedarnos con pos, posiciones de palabras que están contenidas en Voc (el resto se ignora)
- Devolver la estimación MAP





Tema 8. Aprendizaje Bayesiano

Sistemas Inteligentes

de Ciència de la Computació i Intel·ligència *G*rtifici de Ciencia de la Computación e Inteligencia *G*rtifici

Bibliografía

 Mitchell, Machine Learning. McGraw Hill, Computer Science Series. 1997. Capítulo 6



Tema 8. Aprendizaje Bayesiano