Sistemas Probabilísticos

Inteligência Artificial – Prof. Flávio Varejão
Departamento de Informática
Programa de Pós-Graduação em Informática
Universidade Federal do Espírito Santo

Sumário

- Introdução ao Teorema de Bayes
- Sistemas Classificadores Probabilísticos
- Naive Bayes

Teorema de Bayes

- Proposto por Thomas Bayes
- Usado na inferência estatística
- Atualiza estimativas da probabilidade de que diferentes hipóteses sejam verdadeiras
  - Baseia-se nas observações e no conhecimento de como essas observações se relacionam com as hipóteses

Breve Revisão da Estatística

- Probabilidade Condicional de um evento A, tendo ocorrido um evento B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Eventos independentes:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

$$P(A \cap B) = P(A).P(B)$$

Teorema de Bayes

- Teorema de Bayes

$$P(H|E)=\frac{P(E|H).P(H)}{P(E)}$$

- P(H) é a probabilidade *a priori* de H
  - Probabilidade do evento antes que a evidência seja vista
- P(H|E) é a probabilidade *a posteriori* de H
  - Probabilidade do evento depois que a evidência foi vista

Teorema de Bayes

- Dado uma base de exemplos
  - Mais fácil estimar P(E|H) do que P(H|E)
  - Normalmente E = combinação de valores
  - Número de combinações é grande
  - P(E|H) pode ser fatorado em probabilidades mais simples
  - P(H|E) não pode ser fatorado

Sistemas Classificadores Probabilísticos

Naive Bayes

- Em geral, as relações de dependência entre os dados de entrada utilizadas por um classificador são desconhecidas
- Assume-se que os dados são condicionalmente independentes dado a classe
- Naive = “ingênuo”
  - Na maioria dos problemas práticos, essa suposição não é verdade
- Simplifica a abordagem sem comprometer significativamente a precisão do resultado
- Funciona bem na prática

Naive Bayes e a Classificação

- Qual é a probabilidade de uma classe dada uma evidência?
- Naive: evidência é fatorada em partes independentes
  - $P(E|H) = P(E_1|H).P(E_2|H)\dots.\ P(E_n|H)$

$$P(H|E) = \frac{P(E_1|H).P(E_2|H)\dots P(E_n|H).P(H)}{P(E)}$$

Exemplo de Classificação

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Exemplo de Classificação

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Como classificar um novo dia?

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Exemplo de Classificação

Evidência E:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Probabilidade "yes":

$$P(yes|E)=\frac{P(Sunny|yes).P(Cool|yes).P(High|yes).P(True|yes).P(yes)}{P(E)}$$

$$P(yes|E)=\frac{\frac{2}{9}\times\frac{3}{9}\times\frac{3}{9}\times\frac{3}{9}\times\frac{9}{14}}{P(E)}=\frac{0{,}0053}{P(E)}$$

Probabilidade "no":

$$P(no|E)=\frac{P(Sunny|no).P(Cool|no).P(High|no).P(True|no).P(no)}{P(E)}$$

$$P(no|E)=\frac{\frac{5}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{3}{5}\times\frac{5}{14}}{P(E)}=\frac{0{,}0206}{P(E)}$$

Conclusão: "no"

Cálculo das Probabilidades

- A soma das probabilidades de todas as classes é 1
  - $P(yes|E) + P(no|E) = 1$
  - $0{,}0053/P(E) + 0{,}0206/P(E) = 1$
  - $0{,}0053 + 0{,}0206 = P(E)$
  - $P(yes|E) = 0{,}0053/(0{,}0053 + 0{,}0206) = 0{,}2046$
  - $P(no|E) = 0{,}0206/(0{,}0053 + 0{,}0206) = 0{,}7954$

Problema da Freqüência Zero

- O que fazer se um valor de atributo não ocorre em uma classe?
  - Ex: “Humidity = high” para a classe “yes”
  - Probabilidade será zero! (indesejado)
- Solução 1: alguns autores sugerem assumir o valor $1/n^2$ para a probabilidade do atributo que não ocorre, onde n é número de exemplos de treino

Problema da Freqüência Zero

- Solução 2: adicionar 1 unidade a cada combinação de classes e evidências (Laplace estimator).
  - Resultado: Probabilidade nunca será zero.
- Em alguns casos, adicionar uma constante diferente de 1 pode ser mais apropriado.
  - Exemplo: atributo “outlook” para classe “yes”

$$\frac{2+\mu/3}{9+\mu} \qquad \frac{4+\mu/3}{9+\mu} \qquad \frac{3+\mu/3}{9+\mu}$$

Sunny ***Overcast*** ***Rainy***

Os pesos não precisam ser iguais (porém a soma deve ser igual a 1);

$$\frac{2+\mu p_1}{9+\mu} \qquad \frac{4+\mu p_2}{9+\mu} \qquad \frac{3+\mu p_3}{9+\mu}$$

Problema dos Atributos Numéricos

- Nos exemplos anteriores, os atributos eram discretos. O que fazer no caso de atributos contínuos (numéricos)?
- Solução usuais
  - O atributo é discretizado e tratado como um atributo nominal;
  - Ou assumir que os atributos têm uma distribuição de probabilidade *Normal* ou *Gaussiana*.

Problema dos Atributos Numéricos

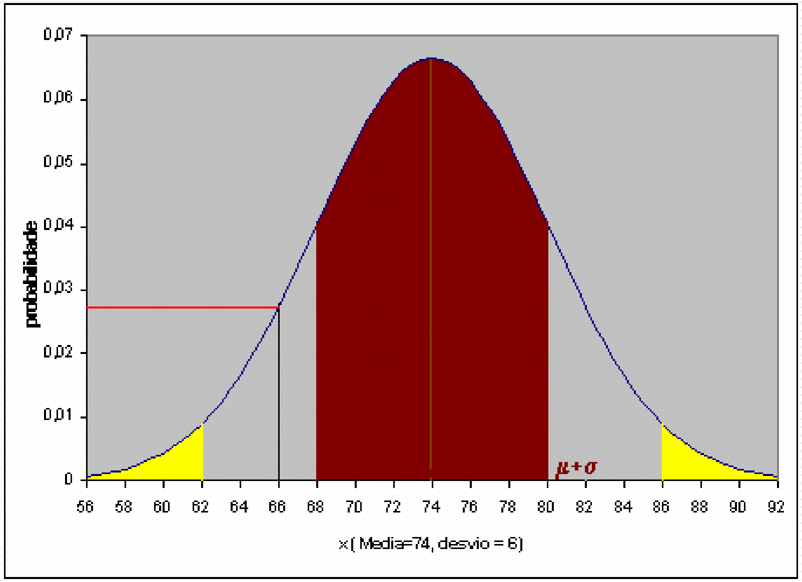
- Solução: Dicretização
  - Quantos intervalos?
    - min(10, número de valores diferentes);
  - Qual amplitude de cada intervalo?
    - Intervalos com mesma amplitude;
    - Intervalos com a mesma freqüência de valores observados;
    - Alternância de classes
    - K-means: K intervalos que minimizem a soma da distância do centro de gravidade de cada intervalo

Problema dos atributos numéricos

- Solução: Assumir distribuição Normal
  - Requer o conhecimento da média (μ) e do desvio padrão (δ) da variável aleatória.

Função densidade de Probabilidade:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Atributos Numéricos

- Atributos contínuos (numéricos):

Outlook			Temperature		Humidity		Windy			Play	
	Yes	*No*	*Yes*	*No*	*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	64, 68,	65, 71,	65, 70,	70, 85,	False	6	2	9	5
Overcast	4	0	69, 70,	72, 80,	70, 75,	90, 91,	True	3	3		
Rainy	3	2	72, …	85, …	80, …	95, …					
Sunny	2/9	3/5	μ =73	μ =75	μ =79	μ =86	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	σ =6.2	σ =7.9	σ =10.2	σ =9.7	True	3/9	3/5		
Rainy	3/9	2/5									

- Exemplo de valor de densidade:

$$f(temperature=66 \mid yes)=\frac{1}{\sqrt{2\pi}6.2}e^{-\frac{(66-73)^2}{2*6.2^2}}=0.0340$$

Problema de Valores Desconhecidos

- Pode ocorrer que algum dos valores dos atributos das entradas de treinamento seja desconhecido
- Solução: Basta não incluir o atributo no cálculo da probabilidade
- Exemplo:

Outlook	Temp.	Humidity	Windy	Play
?	Cool	High	True	?

$$P(class|E) = \frac{P(Cool|class).P(High|class).P(True|class).P(class)}{P(E)}$$

Problema de Valores Desconhecidos

- Outras soluções:
  - No caso de valores nominais:
    - Substituir pela moda;
    - Considerar outro valor possível;
  - No caso de valores contínuos:
    - Substituir pela média;

Naive Bayes: Discussões

- Naive Bayes trabalha surpreendentemente bem, mesmo que a suposição de independência entre os atributos seja claramente violada
- Porque? Classificacão não requer estimativas precisas de probabilidade desde que a maior probabilidade seja atribuída a classe correta
- Entretanto, adicionar muitos atributos redundantes irá causar problemas
- Muitos atributos numéricos não seguem uma distribuição normal