Aprendizagem de Máquina

Aprendizagem Bayesiana

Luiz Eduardo S. Oliveira

Universidade Federal do Paraná
Departamento de Informática
http://web.inf.ufpr.br/luizoliveira

Introdução

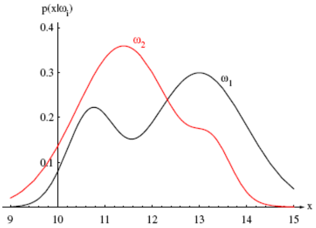
- O pensamento Bayesiano fornece uma abordagem probabilística para a aprendizagem.
- Está baseado na suposição de que as quantidades de interesse são reguladas por distribuições de probabilidades.
- Quantificar o custo/benefício entre diferentes decisões de classificação usando probabilidades e custos associados a classificação.
- Teorema de Bayes
  - Mostra como alterar as probabilidades a priori tendo em conta novas evidências de forma a obter probabilidades a posteriori.

Terminologia

- Classes ω_i (variável aleatória)
- Probabilidades a priori $P(\omega_i)$
  - Conhecimento a priori que se tem sobre o problema, ou seja, conhecimento a priori sobre a aparição de exemplos das classes do problema.
- Função de Densidade Probabilidade $P(x)$
  - Freqüência com a qual encontramos uma determinada característica
  - Evidências

Terminologia

- Densidade de Probabilidade Condicional
  - $P(x|\omega_j)$ (Likelihood)
  - Frequência com que encontramos uma determinada característica dado que a mesma pertence a classe ω_j

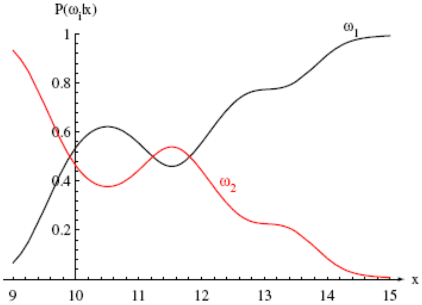


Densidade de duas classes em que x representa uma característica qualquer.

Terminologia

- Probabilidade a posteriori
  - $P(\omega_j|x)$
  - Probabilidade que o padrão pertença a classe ω_j data a característica x
- Regra de decisão baseada apenas em priors
  - ω_1, se $P(\omega_1) > P(\omega_2)$
  - ω_2, caso contrário.

Tomando decisão usando Bayes



Probabilidades a posteriori calculadas usando $P(\omega_1) = 2/3$ e $P(\omega_2) = 1/3$.
Nesse caso, para um valor de $x = 14$, a probabilidade do padrão pertencer a ω_2 é de 0.08, enquanto que a probabilidade do padrão pertencer a ω_1 é de 0.92.
Para cada x, as probabilidades a posteriori somam 1.

$$P(\omega_1|14) = \frac{0.2 \times 2/3}{0.2 \times 2/3 + 0.04 \times 1/3} = 0.92$$

Teorema de Bayes

- Basicamente o teorema de Bayes mostra como rever as crenças sempre que novas evidências são coletadas.
- Ou seja, atualizar a probabilidade a posteriori utilizando para isso a probabilidade a priori e as verossimilhanças e as evidências

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

- $P(A|B)$ é a probabilidade a posteriori
- $P(A)$ é a probabilidade a priori
- $P(B|A)$ são as verossimilhanças (likelihood)
- $P(B)$ são as evidências, dado por $\sum P(A_i) \times P(B|A_i)$

Exemplo

- Um médico sabe que a meningite causa torcicolo em 50% dos casos. Porém, o médico sabe que a meningite atinge 1/50000 e também que a probabilidade de se ter torcicolo é de 1/20.
- Usando Bayes pra saber a probabilidade de uma pessoa ter meningite dado que ela está com torcicolo

Exemplo

- Um médico sabe que a meningite causa torcicolo em 50% dos casos. Porém, o médico sabe que a meningite atinge 1/50000 e também que a probabilidade de se ter torcicolo é de 1/20.
- Usando Bayes pra saber a probabilidade de uma pessoa ter meningite dado que ela está com torcicolo

Temos então

- $P(T|M) = 0.5$
- $P(M) = 1/50000$
- $P(T) = 1/20$

$$P(M|T) = \frac{P(M) \times P(T|M)}{P(T)} = \frac{1/50000 \times 0.5}{1/20} = 0.0002$$

Exercício

- Considere o sistema de classificação de peixes. Para essa época do ano, sabe-se que a probabilidade de pescar salmão é maior que pescar robalo, $P(salmao) = 0.82$ e $P(robabo) = 0.18$.
- Suponha que a única característica que você pode contar é a intensidade do peixe ou seja, se ele é claro ou escuro. Sabe-se que 49.5% dos salmões tem intensidade clara e que 85% dos robalos tem intensidade clara.
- Calcule a probabilidade de ser salmão dado que o peixe pescado tem intensidade clara.

Exercício

- Considere o sistema de classificação de peixes. Para essa época do ano, sabe-se que a probabilidade de pescar salmão é maior que pescar robalo, $P(salmao) = 0.82$ e $P(robabo) = 0.18$.
- Suponha que a única característica que você pode contar é a intensidade do peixe ou seja, se ele é claro ou escuro. Sabe-se que 49.5% dos salmões tem intensidade clara e que 85% dos robalos tem intensidade clara.
- Calcule a probabilidade de ser salmão dado que o peixe pescado tem intensidade clara.

$$P(S|C) = \frac{P(S) \times P(C|S)}{P(C)} = \frac{0.82 \times 0.495}{0.82 \times 0.495 + 0.18 \times 0.85} = 0.726$$

Classificador Naïve Bayes

- Um dos algoritmos de aprendizagem mais práticos e utilizados na literatura.
- Denominado Naïve (ingênuo) por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não é informativa sobre nenhum outro.
- Apesar dessa premissa, o classificador reporta bom desempenho em diversas tarefas de classificação.
- Aplicações bem sucedidas:
  - Diagnóstico
  - Classificação de documentos textuais

Classificador Naïve Bayes

- Se aplica a tarefas de aprendizagem onde cada instância x é descrita por um conjunção de valores de atributos em que a função alvo, $f(x)$ pode assumir qualquer valor de um conjunto V.
- Um conjunto de exemplos de treinamento da função alvo é fornecido a uma nova instância é apresentada, descrita pela tupla de valores de atributos $< a_1, a_2, \ldots, a_n >$.
- A tarefa é predizer o valor alvo (ou classificação) para esta nova instância.

Classificador Naïve Bayes

- O classificador é baseado na suposição de que os valores dos atributos são condicionalmente independentes dados o valor alvo.
- Se usarmos Bayes para múltiplas evidências, temos

$$P(H|E_1, E_2, \ldots, E_n) = \frac{P(E_1, E_2, \ldots, E_n|H) \times P(H)}{P(E_1, E_2, \ldots, E_n)}$$

- Considerando a hipótese de independência, podemos re-escrever o teorema de Bayes da seguinte forma:

$$P(H|E_1, E_2, \ldots, E_n) = \frac{P(E_1|H)\times, P(E_2|H)\times, \ldots, P(E_n|H) \times P(H)}{P(E_1, E_2, \ldots, E_n)}$$

O denominador pode ser ignorado por se tratar de um termo comum.

Exemplo

Para entender melhor o considere o seguinte problema:

	outlook	temperature	humidity	windy	play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

Construindo o Modelo (NB)

O primeiro passo consiste em construir o modelo de probabilidades condicionais Naïve Bayes (NB)

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								

- A tabela acima contem a frequência de diferentes evidências.
- Por exemplo, existem duas instâncias mostrando (outlook=sunny) para (jogar=yes)

Avaliação

Após definir todas as frequências é necessário calcular todas as probabilidades condicionais e as probabilidades a priori.

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
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								

Por exemplo:

- $P(\text{outlook}=\text{sunny}|\text{play}=\text{yes}) = 2/9$
- $P(\text{play}=\text{yes}) = 9/14$

Predição

- De posse do modelo, podemos usá-lo para predizer um evento "play" com base em um conjunto qualquer de evidências.
- Por exemplo: `[Sunny,Cool,High,True,?]`

$$
\begin{aligned}
P(Yes|E) \quad = \quad & (P(Outlook = sunny|Yes) \times \\
& P(Temp = Cool|Yes) \times \\
& P(Humidity = High|Yes) \times \\
& P(Windy = True|Yes) \times \\
& P(Yes))/P(E)
\end{aligned}
$$

P(E) pode ser ignorada por se tratar de um denominador comum quando queremos comparar as duas classes. Deste modo, temos

$$
P(Yes|E) = \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}
$$

Predição

Calculando a predição para as duas classes

- Para "Yes" temos $\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} = 0.0053$
- Para "No" temos $\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14} = 0.0206$

Convertendo esses valores para probabilidade através da normalização, temos

- $P(Yes|E) = 0.0053/(0.0053 + 0.0206) = 0.205$
- $P(No|E) = 0.0206/(0.0053 + 0.0206) = 0.795$

Técnica de Suavização

- Em alguns casos, a frequência pode ser zero, como por exemplo P(outlook=overcast|play=No) = 0/5.
- Isso cria um problema para calcular P(No), a qual será sempre zero quando esta evidência for utilizada.
- A técnica de suavização mais utilizada é a estimação de Laplace, a qual é dada

$$P'(H|E) = \frac{n_c + \mu p}{n + \mu}$$

- n_c é o número de hipóteses existentes para a classe (Ex: Zero para outlook=overcast e play=no)
- n número e exemplos totais para o treinamento
- μ número de exemplos virtuais
- Considerado que as evidências são igualmente distribuídas, tempo $p = \frac{1}{3}$ (sunny,overcast,rainy)

Técnica de Suavização

Reestimando os valores usando Laplace, teriamos

- $P(outlook = Sunny|play = No) = \frac{3+3\times1/3}{5+3} = \frac{4}{8}$
- $P(outlook = Overcast|play = No) = \frac{0+3\times1/3}{5+3} = \frac{1}{8}$
- $P(outlook = Rainy|play = No) = \frac{2+3\times1/3}{5+3} = \frac{3}{8}$

Desta forma, todos os valores foram redistribuídos mantendo uma proporção similar

Calculando as probabilidade para atributos contínuos

Existem duas maneiras

- Discretizar os atributos continuos em algumas categorias. Por exemplo, temperatura acima de 80F pode ser considerada alta.
- Outra forma consiste em usar uma função de densidade probabilidade e desta forma preservar os valores continuos.
  - Nesse caso assumimos que as variáveis continuas seguem uma distribuição normal
  - Com isso em mente, podemos calcular a média e desvio de cada variável usando a base de aprendizagem.
  - De posse da média e desvio, basta aplicar a formula da normal para estimar a probabilidade

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

- scipy.stats.norm$(\mu, \sigma).pdf(x)$

Exemplo

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

Normal distribution

		Humidity	Mean	StDev
Play Golf	yes	86 96 80 65 70 80 70 90 75	79.1	10.2
	no	85 90 70 95 91	86.2	9.7

$$P(\text{humidity} = 74 \mid \text{play} = \text{yes}) = \frac{1}{\sqrt{2\pi}(10.2)} e^{-\frac{(74-79.1)^2}{2(10.2)^2}} = 0.0344$$

$$P(\text{humidity} = 74 \mid \text{play} = \text{no}) = \frac{1}{\sqrt{2\pi}(9.7)} e^{-\frac{(74-86.2)^2}{2(9.7)^2}} = 0.0187$$

Exercício

Attribute	Class yes (0.63)	no (0.38)
outlook		
sunny	3.0	4.0
overcast	5.0	1.0
rainy	4.0	3.0
[total]	12.0	8.0
temperature		
mean	72.9697	74.8364
std. dev.	5.2304	7.384
weight sum	9	5
precision	1.9091	1.9091
humidity		
mean	78.8395	86.1111
std. dev.	9.8023	9.2424
weight sum	9	5
precision	3.4444	3.4444
windy		
TRUE	4.0	4.0
FALSE	7.0	3.0
[total]	11.0	7.0

Calcular a probabilidade para

E =[Outlook=rainy, Temp=65, Humid, 70, Wind=True)

$P(Yes|E) = ?$

$P(No|E) = ?$