

Aprendizado de Máquina

Métodos probabilísticos

Prof. Dr. André C. P. L. F. de Carvalho
ICMC-USP



Exemplo

- Predizer se uma pessoa terá problemas cardíacos
 - Atributos preditivos: peso e frequência de exercício
 - Ignora outras possíveis causas:
 - Bebida
 - Hereditariedade
 - Fumo
 - Stress
 - ...

24/05/2019

André de Carvalho - ICMC/USP

4

Principais tópicos

- Métodos baseados em probabilidade
- Métodos discriminativos
 - Regressão Logística
- Métodos generativos
 - Teoria das probabilidades
 - Teorema de Bayes
 - *Naive Bayes*

24/05/2019

André de Carvalho - ICMC/USP

2

Métodos probabilísticos

- Em várias aplicações é importante ...
 - Estimar a probabilidade de um exemplo pertencer a uma classe
- Modelam relacionamento probabilístico entre atributos preditivos e atributo alvo
- Tipos de modelos induzidos:
 - Modelos discriminativos
 - Modelos generativos

24/05/2019

André de Carvalho - ICMC/USP

5

Introdução

- Muitos problemas de classificação são não determinísticos
 - Relação entre atributos de entrada e classe é probabilística
 - Ruído nos dados
 - Algumas informações importantes não são capturadas pelos atributos preditivos usados
 - Informações capturadas pelos atributos preditivos usados são incompletas ou imprecisas

24/05/2019

André de Carvalho - ICMC/USP

3

Métodos discriminativos

- Modelam a distribuição de probabilidade a posteriori (condicional) $P(Y/X)$
- Dado X , retornam a probabilidade de Y ocorrer
 - X : atributo(s) preditivo(s)
 - Y : atributo alvo
 - Ex.: Regressão logística

24/05/2019

André de Carvalho - ICMC/USP

6

Métodos generativos

- Modelam a distribuição de probabilidade conjunta $P(X,Y)$
 - Com a distribuição conjunta é possível derivar qualquer distribuição condicional
- Induzidos por algoritmos baseados no teorema de Bayes
 - Algoritmos Bayesianos
 - Ex.: *Naive Bayes*

24/05/2019 André de Carvalho - ICMC/USP 7

Regressão logística

- Apesar do nome, é usada para tarefas de classificação
 - Regressão: regressão linear
- Estima probabilidade que um exemplo pertence a uma dada classe
 - Ajusta uma função logística a um conjunto de dados
 - Utiliza um conjunto de treinamento
 - Gera uma curve sigmoide
 - Produz um hiperplano de separação

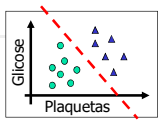


24/05/2019 André de Carvalho - ICMC/USP 10

Discriminante linear

- Induz função linear
 - Função discriminante
 - Ajusta parâmetros da função discriminante

$$f(x) = w_0 + w_1x_1 + w_2x_2 + \dots$$
 - Valor de $f(x)$
 - Distância de x à fronteira
 - Estima chance de x pertencer à classe pos. (neg.)
 - Ex.: círculo₁ está a $-m$ e triângulo₁ está a $+k$ da reta
 - Semelhante à rede Perceptron

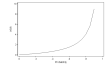


24/05/2019 André de Carvalho - ICMC/USP 8

Regressão logística

P_+ : Probabilidade de sucesso	Probabilidade (P_+)	Chance ($P_+/(1-P_+)$)	Log(Chance)
	0,5	50:50 (1:1) = 1	0,00
	0,9	90:10 (9:1) = 9	2,19
	0,999	999:1 = 999	6,91
	0,01	1:99 = 0,0101	-4,60
	0,001	1:999 = 0,001001	-6,91

- Encontrar $f(x)$ que modele $\log(\text{Chance})$
 - Permite estimar probabilidade usando modelo gerado por discriminante linear



© André de Carvalho - ICMC/USP 11

Discriminante linear

- Distância de exemplos a fronteira de decisão definida por uma função linear
- Problema:
 - Distância: $-\infty < f(x) < +\infty$
 - Modelos probabilísticos devem retornar uma probabilidade: $0 < f(x) < 1$
- Solução:
 - Regressão logística

© André de Carvalho - ICMC/USP 9

Regressão logística

- Probabilidade de exemplo pertencer a classe positiva
 - Evento ocorreu

$$\text{Função logit} \rightarrow \log\left(\frac{p_+(x)}{1-p_+(x)}\right) = f(x) = w_0 + w_1x_1 + w_2x_2 + \dots$$

$$p_+(x) = \frac{1}{1+e^{-f(x)}}$$

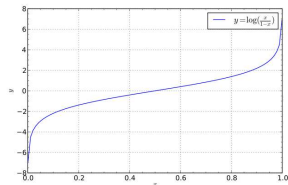
- Treinamento

$$g(x, w) = \begin{cases} p_+(x) & \text{se } x \text{ é } + \\ 1-p_+(x) & \text{se } x \text{ é } - \end{cases}$$
 Função objetivo para ajuste dos pesos

© André de Carvalho - ICMC/USP 12

Regressão logística

- Função logit
 - Inversa da função logística



$$\log\left(\frac{p_+(x)}{1-p_+(x)}\right)$$

24/05/2019

André de Carvalho - ICMC/USP

13

Teoria das probabilidades

- $P(A)$ satisfaz axiomas de Kolmogorov
 - $P(A) \geq 0$
 - $P(\Omega) = 1$
 - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 - Se A e B são eventos mutuamente exclusivos
 - $(A \cap B) = \emptyset$
 - $P(A \cap B) = 0$
 - $P(A \cup B) = P(A) + P(B)$

24/05/2019

André de Carvalho - ICMC/USP

16

Treinamento

- Encontrar valores de w_i que minimizem erro no conjunto de treinamento
 - Faz aproximação numérica utilizando método de máxima verossimilhança
 - Gradiente descendente estocástica
 - Para grandes conjuntos de dados
 - Exemplo para 1 atributo preditivo
 - w_0 : posição da função logística
 - w_1 : inclinação da função logística

24/05/2019

André de Carvalho - ICMC/USP

14

Probabilidade conjunta

- Probabilidade conjunta
 - Probabilidade de dois eventos ocorrerem simultaneamente
 - $P(A \cap B)$ ou $P(A, B)$
 - Se eventos são eventos independentes
 - A ocorrência de um não afeta a probabilidade de ocorrência do outro
 - $P(A \cap B) = P(A) * P(B)$

24/05/2019

André de Carvalho - ICMC/USP

17

Teoria das probabilidades

- Espaço amostral (Ω): todas as possíveis observações de um experimento
- Evento (A): subconjunto de possíveis observações em Ω
- Ex.: Valores de um dado de 6 faces
 - $\Omega = \{1, 2, 3, 4, 5, 6\}$
 - A = valor do dado $< 3 = \{1, 2\}$
 - A = valor do dado é par = $\{2, 4, 6\}$
 - $P(A)$: probabilidade de um evento ocorrer

24/05/2019

André de Carvalho - ICMC/USP

15

Probabilidade e AM

- Tarefa: dado o resultado de um exame, prever se paciente está doente
- Atributo preditivo
 - Resultado de exame
- Atributo alvo
 - Diagnóstico do paciente
 - Predição (classificação)

© André de Carvalho - ICMC/USP

18

Probabilidade e AM

- Sejam dois eventos A e B
 - A: atributo alvo (presença de uma doença)
 - Variável aleatória com dois valores: presença e ausência
 - B: atributo de entrada (resultado de um exame)
 - Variável aleatória com dois valores: positivo e negativo
 - P(A): probabilidade do evento A ocorrer (presença da doença)
 - $P(A) = 1 - P(\neg A)$
 - P(B): probabilidade do evento B ocorrer (exame positivo)
 - $P(B) = 1 - P(\neg B)$

24/05/2019 André de Carvalho - ICMC/USP 19

Probabilidade e AM

Paciente	Exame	Doença
001	positivo	presente
002	negativo	presente
003	negativo	ausente
004	positivo	presente
005	positivo	ausente
006	positivo	presente
007	negativo	ausente
008	negativo	presente
009	positivo	ausente
010	positivo	presente

Probabilidade da variável preditiva e probabilidade *a priori* podem ser estimadas pela frequência

$P(\text{negativo}) = 0,4$
 $P(\text{positivo}) = 0,6$
 $P(\text{presente}) = 0,6$
 $P(\text{ausente}) = 0,4$

O que se deseja em AM é a probabilidade *a posteriori*

24/05/2019 André de Carvalho - ICMC/USP 22

Probabilidade e AM

- P(A): probabilidade a priori do paciente está doente
- P(B): distribuição da variável preditiva B ser verdade (exame deu resultado positivo)
 - Evidência
- P(B/A): probabilidade de verosimilhança
 - Para um valor fixo de B, define verosimilhança (plausibilidade) de cada um dos possíveis valores de A

© André de Carvalho - ICMC/USP 20

Probabilidade condicional

- Probabilidade de ocorrência de um evento dada a ocorrência de outro
 - P(A/B)
 - Probabilidade de ocorrência de um evento A dada a ocorrência de um evento B
 - Ex.: Probabilidade de estar doente (A) dado que um exame (B) deu positivo
 - Atributos (eventos) independentes: $P(A/B) = P(A)$

24/05/2019 André de Carvalho - ICMC/USP 23

Probabilidade e AM

Paciente	Exame	Doença
001	positivo	presente
002	negativo	presente
003	negativo	ausente
004	positivo	presente
005	positivo	ausente
006	positivo	presente
007	negativo	ausente
008	negativo	presente
009	positivo	ausente
010	positivo	presente

Probabilidade da variável preditiva e probabilidade *a priori* podem ser estimadas pela frequência

$P(\text{negativo}) =$
 $P(\text{positivo}) =$
 $P(\text{presente}) =$
 $P(\text{ausente}) =$

O que se deseja em AM é a probabilidade *a posteriori*

24/05/2019 André de Carvalho - ICMC/USP 21

Probabilidade condicional

- Fácil estimar pela frequência as probabilidades *a priori*
 - P(B): prob. do resultado do exame ser positivo
 - P(A): prob. do paciente estar doente
 - P(B/A): prob. do resultado do exame ser positivo dado que o paciente esta doente
- Difícil estimar probabilidade *a posteriori*
 - P(A/B): probabilidade do paciente estar doente dado que seu exame deu positivo
 - Teorema (regra) de Bayes

24/05/2019 André de Carvalho - ICMC/USP 24

Probabilidade condicional

- Lei da probabilidade condicional
 - $P(A/B) = P(A \cap B) / P(B)$
- Teorema de Bayes
 - Permite calcular probabilidade *a posteriori* de um evento
 - $P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$
 - $P(A/B) = P(B/A)P(A)/P(B)$
 - Posteriori* = (verossimilhança x *priori*) / *evidência*
 - $P(B)$: lei da probabilidade total

24/05/2019 André de Carvalho - ICMC/USP 25

Classificação Bayesiana

- Sejam $y_i, i = 1, 2, \dots, m$, as possíveis classes
 - Novo exemplo pertence à classe que tiver probabilidade *a posteriori* máxima
 - $Y_{MAP} = \arg \max_i P(y_i/X)$
- Definição de $P(y_i/X)$
 - $P(y_i/X) = P(X/y_i) P(y_i) / P(X)$

24/05/2019 Redes Neurais - André Ponce de Leon F. de Carvalho - IABIC/USP 28

Probabilidade condicional

- Lei da probabilidade total
 - Evento A pode ter 2 possíveis resultados, A_1 e $\neg A_1$ (A_2), que formam uma partição em Ω

$$P(B) = P(B \cap A_1) + P(B \cap A_2)$$

$$P(B) = P(B / A_1)P(A_1) + P(B / A_2)P(A_2)$$
 - Evento A pode ter n possíveis resultados mutuamente exclusivos, A_1, A_2, \dots, A_n , que formam uma partição em Ω

$$P(B) = \sum_{i=1}^n P(B / A_i)P(A_i)$$

24/05/2019 André de Carvalho - ICMC/USP 26

Classificação Bayesiana

- Expressão $P(X/y_i) P(y_i) / P(X)$ pode ser simplificada
 - $P(X)$ é comum a todas as classes
 - Considerar as classes equiprováveis ($P(y_i) = P(y_j)$)
- Exemplo x pertence a classe com máxima verossimilhança
 - $h_{MV} = \arg \max_i P(X/y_i)$
- Difícil calcular valores
 - Precisa de um número de exemplos muito grande

24/05/2019 André de Carvalho - ICMC/USP 29

Probabilidade condicional

De acordo com experiências passadas

Doença	Exame	
	positivo	negativo
presente: 8%	75%	4%
ausente: 92%	25%	96%

$P(\text{Exame}/\text{Doença}) = 0,75$
 $P(\neg \text{Exame} / \neg \text{Doença}) = 0,96$
 $P(\text{Exame}) = P(\text{Exame}/\text{Doença})P(\text{Doença}) + P(\text{Exame}/\neg \text{Doença})P(\neg \text{Doença})$
 $P(\text{Exame}) = 0,75 \times 0,08 + 0,25 \times 0,92 = 0,29$
 ...
 $P(\text{Doença}/\text{Exame}) = ?$

24/05/2019 André de Carvalho - ICMC/USP 27

Classificação Bayesiana

- Inferência Bayesiana
 - Cálculo da probabilidade *a posteriori* a partir da probabilidade *a priori*
- Várias alternativas para estimar $P(X/y_i)$
 - Produzem diferentes funções discriminantes
 - Ex.: Classificador Naive Bayes

24/05/2019 André de Carvalho - ICMC/USP 30

Naive Bayes

- Classificador Bayesiano mais simples
- Assume que os atributos são independentes
 - $P(X/y_i) = P(x_1/y_i) * \dots * P(x_d/y_i)$

$$P(y_i / X) \propto P(y_i) \prod_{j=1}^d P(x_j / y_i)$$

$$\log P(y_i / X) \propto \log P(y_i) + \sum_{j=1}^d \log P(x_j / y_i)$$

24/05/2019

André de Carvalho - ICMC/USP

31

Exercício

- Seja o seguinte cadastro de pacientes:

Nome	Febre	Enjôo	Manchas	Dores	Diagnóstico
João	sim	sim	pequenas	sim	doente
Pedro	não	não	grandes	não	saudável
Maria	não	sim	pequenas	não	saudável
José	sim	sim	grandes	sim	doente
Ana	sim	não	pequenas	sim	saudável
Leila	não	não	grandes	sim	doente

© André de Carvalho - ICMC/USP

34

Naive Bayes

- Para duas classes

$$\log \frac{P(y_1 / X)}{P(y_2 / X)} \propto \log \frac{P(y_1)}{P(y_2)} + \sum_{j=1}^d \log \frac{P(x_j / y_1)}{P(x_j / y_2)}$$

- Sinal do primeiro log indica a classe
- Sinal de cada termo do somatório indica contribuição de cada atributo

24/05/2019

André de Carvalho - ICMC/USP

32

Exercício

- Utilizar Naive Bayes para induzir modelo capaz de distinguir:
 - Pacientes potencialmente saudáveis
 - Pacientes potencialmente doentes
- Testar o modelo para novos casos
 - (Luis, não, não, pequenas, sim)
 - (Laura, sim, sim, grandes, sim)

© André de Carvalho - ICMC/USP

35

Exercício

- Usar Naive Bayes para gerar um modelo probabilístico
 - Usar exemplos de treinamento (001, -1) e (110, +1)
 - Definir a classe dos exemplos: 111, 000, 100 e 011

© André de Carvalho - ICMC/USP

33

Conclusão

- Métodos baseados em probabilidade
- Teorema de Bayes
- Naive Bayes
- Classificadores Bayesianos
- Classificadores Bayesianos com k dependências

24/05/2019

André de Carvalho - ICMC/USP

36



Perguntas

