Data Science Academy

www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Inferência em Redes Bayesianas

A inferência bayesiana é um tipo de inferência estatística que descreve as incertezas sobre quantidades invisíveis de forma probabilística. Incertezas são modificadas periodicamente após observações de novos dados ou resultados. A operação que calibra a medida das incertezas é conhecida como operação bayesiana e é baseada na fórmula de Bayes. A fórmula de Bayes é muitas vezes denominada Teorema de Bayes.

Inferência bayesiana deriva a probabilidade posterior como consequência de dois antecedentes, uma probabilidade anterior e uma "função de verossimilhança" derivado de um modelo de probabilidade para os dados a serem observados. A inferência bayesiana calcula a probabilidade posterior de acordo com a regra de Bayes:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

onde:

- $\mid$ significa *dado*.
- H representa qualquer *hipótese* cuja probabilidade pode ser afetado por dados (chamada evidência abaixo). Muitas vezes, há hipóteses concorrentes , a partir do qual se escolhe a mais provável.
- a evidência E, corresponde a novos dados que não foram utilizados no cálculo da probabilidade anterior.
- $P(H)$, a *probabilidade a priori*, é a probabilidade de H antes de E ser observada. Isso indica sua estimativa anterior da probabilidade de que uma hipótese seja verdadeira, antes de obter a evidência atual.
- $P(H \mid E)$ a *probabilidade a posteriori*, é a probabilidade de H dado E, ou seja , *depois* de E ser observada. Isto diz-nos o que queremos saber: a probabilidade de uma hipótese *dada* a evidência observada.
- $P(E \mid H)$ é a probabilidade de observar E *dado* H. indica a compatibilidade dos elementos com a hipótese dada.
- $P(E)$ é às vezes chamado de probabilidade marginal ou "evidência modelo" . Este fator é a mesma para todas as hipóteses possíveis que estão sendo considerados. Isto significa que este fator não entra em determinar as probabilidades relativas de diferentes hipóteses .

A tarefa básica para qualquer sistema de inferência probabilístico é calcular a distribuição de probabilidade posterior para um conjunto de variáveis de consulta, dado algum evento observado, isto é, alguma atribuição de valores a um conjunto de variáveis de evidência. Na rede de alarme contra roubo, poderíamos observar o evento em que JoãoLiga = verdadeiro e MariaLiga = verdadeiro. Então, poderíamos buscar, digamos, a probabilidade de ter ocorrido um roubo:

$$\mathbf{P}(Roubo \mid JoãoLiga = verdadeiro, MariaLiga = verdadeiro) = \langle 0{,}284,\ 0{,}716 \rangle$$

Certamente, entende-se que a regra de bayes faz muito sentido. Se a evidência não corresponder com a hipótese, a hipótese deve ser rejeitada. Mas se uma hipótese é extremamente improvável a priori, deve-se também rejeitá-la, mesmo que a evidência pareça corresponder-se.

Por exemplo, imagine que eu tenho várias hipóteses sobre a natureza de um bebê recém-nascido, incluindo:

- É um menino de cabelo marrom.
- É uma menina de cabelos loiros.
- É um cachorro.

Então considere dois cenários:

1. Estou apresentado com provas sob a forma de uma imagem de uma menina de cabelos loiros. Acho que esta evidência apoia e se opõe a de que é um menino de cabelo marrom.
2. Estou apresentado com provas sob a forma de uma imagem de um cão. Apesar desta evidência, tratada isoladamente, apoio, a minha crença anterior nesta hipótese (que um ser humano pode dar à luz a um cachorro) é extremamente pequena, por isso a probabilidade posterior é, no entanto, desprezível.

O ponto crítico sobre inferência Bayesiana, então, é que ele fornece uma maneira de princípios de combinar novas evidências com as crenças anteriores, através da aplicação da regra de Bayes. Além disso, a regra de Bayes pode ser aplicada de forma iterativa: depois de observar algumas provas, a probabilidade posterior resultante pode então ser tratado como uma probabilidade prévia, e uma nova probabilidade posterior calculado a partir de novas provas. Isto permite à princípios Bayesianos serem aplicado a vários tipos de evidência, se visto de uma só vez ou ao longo do tempo. Este procedimento é chamado de "atualização bayesiana".

Principais Algoritmos Usados em Inferência Bayesiana:

Inferência Exata em Redes Bayesianas

1- Inferência por enumeração

função ASK-ENUMERAÇÃO(X, **e**, rb) **retorna** uma distribuição sobre X
entradas: X, a variável de consulta
e, valores observados para variáveis **E**
rb, uma rede bayesiana com variáveis $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *variáveis ocultas* */

Q(X) ← uma distribuição sobre X, inicialmente vazia
para cada valor x_i de X **faça**
estender **e** com valor xi para X
Q(xi) ← ENUMERAR-TODOS(bn.VARS, e_{xi})
Where $\mathbf{e}_{xi}$ é **e** estendido com $X = x_i$
retornar NORMALIZAR(**Q**(X))

---

função ENUMERAR-TODOS($vars$, **e**) **retorna** um número real
se VAZIO?($vars$) **então retornar** 1,0

Y ← PRIMEIRO($vars$)
se Y tem valor y em **e**
então retornar $P(y \mid pais(Y))$ × ENUMERAR-TODOS(RESTO($vars$), **e**)
senão retornar Σ_y P(y | pais(y)) × ENUMERAR-TODOS (vars), e_y)
onde $\mathbf{e}_{yi}$ é **e** estendido com $Y = y_i$

2- Inferência por Eliminação de Variáveis

função ASK-ELIMINAÇÃO(X, **e**, rb) **retorna** uma distribuição sobre X
entradas: X, a variável de consulta
e, variáveis observadas da variável **E**
rb, uma rede bayesiana especificando a distribuição conjunta $\mathbf{P}(X_1, ..., X_n)$

fatores ← [ ]
para cada *var* **em** ORDEM(*rb*.VARS) **faça**
fatores ← [CRIAR-FATOR (*var*, **e**) *fatores*]
se *var* é uma variável oculta **então** *fatores* ← SOMAR(*var* | *fatores*)
retornar NORMALIZAR(PRODUTO-PONTUAL(*fatores*))

Inferência Aproximada em Redes Bayesianas

Dada a intratabilidade da inferência exata em redes extensas com várias conexões, é essencial considerar métodos de inferência aproximada. Esses algoritmos de amostragem aleatória, também chamados algoritmos de Monte Carlo, fornecem respostas aproximadas cuja exatidão depende do número de amostras geradas. Em anos recentes, os algoritmos de Monte Carlo dos quais a têmpera simulada (que estudamos no capítulo 2) é um exemplo, são utilizados em muitas ramificações da ciência da computação para estimar quantidades que são difíceis de calcular com exatidão. Principais algoritmos para inferência aproximada em redes bayesianas:

1. Amostra a priori

função AMOSTRA-A-PRIORI(*rb*) **retorna** um evento amostrado a partir da probabilidade *a priori* especificada por *rb*
  entradas: *rb*, uma rede bayesiana que especifica a distribuição conjunta $\mathbf{P}(X_1, ..., X_n)$

  x ← um evento com *n* elementos
  para cada variável X_i, **em** $X_1,....,X_n$ **faça**
    x[*i*] ← *uma amostra aleatória de* $\mathbf{P}(X_i \mid pais(X_i))$
  retornar x

2. Amostragem de rejeição

função AMOSTRAGEM-DE-REJEIÇÃO(*X*, **e**, *rb*, *N*) **retorna** uma estimativa de $\mathbf{P}(X \mid \mathbf{e})$
  entradas: *X*, a variável de consulta
    e, valores observados para variáveis **E**
    rb, uma rede bayesiana
    N, o número total de amostras a serem geradas
  variáveis locais: N, um vetor de contagens para cada valor de *X*, inicialmente zero
  para $j = 1$ **até** *N* **faça**
    x ← AMOSTRA-A-PRIORI(*rb*)
    se x é consistente com **e então**
      N[*x*] ← **N**[*x*]+1 onde *x* é o valor de *X* em **x**
  retornar NORMALIZAR(**N**)

3. Ponderação de probabilidade

função PONDERAÇÃO-DE-PROBABILIDADE (X, **e**, rb, N) **retorna** uma estimativa de $\mathbf{P}(X \mid \mathbf{e})$
  entradas: X, a variável de consulta
    e, valores observados para variáveis **E**
    rb, uma rede bayesiana especificando distribuição conjunta $\mathbf{P}(X_1,\ldots,X_n)$
    N, o número total de amostras a serem geradas
  variáveis locais: **W**, um vetor de contagens ponderadas para cada valor de X, inicialmente igual a
  para $j = 1$ **até** N **faça**
    x, $w \leftarrow$ AMOSTRA-PONDERADA(rb, e)
    $\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ onde x é o valor de X em **x**
  retornar NORMALIZAR(**W**)

---

função AMOSTRA-PONDERADA(rb, **e**) **retorna** um evento e um peso
  $w \leftarrow 1$; $\mathbf{x} \leftarrow$ um evento com n elementos inicializados de **e**
  para cada variável X_i **em** $X_1,\ldots,X_n$ **faça**
    se X_i é uma variável de evidência com valor x_i em **e**
      então $w \leftarrow w \times P(X_i = x_i \mid pais(X_i))$
      else $\mathbf{x}[i] \leftarrow$ uma amostra aleatória de $\mathbf{P}(X_i \mid \text{pais}(X_i))$
  retornar **x**, w

Veremos a implementação de alguns desses algoritmos em Python, na sequência do treinamento.

Referências:

Livro: Inteligência Artificial
Autor: Peter Norvig