Rastreio de Processo Bayesiano

Manoel Galdino

2024-05-28

Regra do Produto

Definição

Para dois eventos (independentes ou não) A e B, a probabilidade conjunta $p(AB)$ é dada por:

$$p(A) \cdot p(B|A)$$

e, igualmente:

$$p(B) \cdot p(A|B)$$

Similarmente, para trÊs eventos A, B e C, a probabilidade conjunta $p(ABC)$ pode ser decomposta em uma série de probabilidades condicionais:

$$p(ABC) = p(A) \cdot p(B|A) \cdot p(C|AB)$$

Explicação

1. **Probabilidade de** A: Primeiro, consideramos a probabilidade do evento A.
2. **Probabilidade de** B **dado** A: Em seguida, calculamos a probabilidade do evento B dado que A já ocorreu.
3. **Probabilidade de** C **dado** A **e** B: Finalmente, determinamos a probabilidade do evento C dado que A e B já ocorreram.

Teorema de Bayes

Se

$$p(DH) = p(H) \cdot p(D|H) = p(D) \cdot p(H|D)$$

, então:

$$p(H|D) = \frac{p(H) \cdot p(D|H)}{p(D)}$$

Chamamos $P(H)$ de priori de A, $p(D|H)$ de verossimilhança e $p(D)$ de priori de D ou constante normalizadora.

Constante normalizadora

- Aplicando a regra da probabilidade total, podemos calcular $P(D)$:
- $P(D) = p(H) \cdot p(D|H) + p(\neg H) \cdot p(D|\neg H)$
- Ou, podemos estimar a *posterior odds*:
  $\frac{p(H|D)}{p(\neg H|D)} = \frac{\frac{p(H) \cdot p(D|H)}{p(D)}}{\frac{p(\neg H) \cdot p(D|\neg H)}{p(D)}} = \frac{p(H) \cdot p(D|H)}{p(\neg H) \cdot p(D|\neg H)}$

Função de Verossimilhança

Definição Verbal

A função de verossimilhança é a probabilidade dos dados observados tratada como uma função de parâmetros.

Interpretação da Função de Verossimilhança

Podemos escrever $p(x_1, x_2, \ldots, x_n|\theta)$ de duas formas distintas:

1. **Distribuição de Probabilidade Conjunta para os Dados**
   - Dado θ, temos uma distribuição de probabilidade conjunta para observar certos valores dos n dados.
   - Aqui, os dados são aleatórios e θ é fixo.

$$p(x_1, x_2, \ldots, x_n|\theta)$$

2. **Função do Parâmetro: Verossimilhança**
   - Os dados são fixos (já coletados) e o parâmetro θ é aleatório.
   - A verossimilhança é tratada como uma função do parâmetro θ.

$$L(\theta|x_1, x_2, \ldots, x_n) = p(x_1, x_2, \ldots, x_n|\theta)$$

Distribuição de Probabilidade Conjunta

Explicação

- Quando interpretamos $p(x_1, x_2, \ldots, x_n|\theta)$ como uma distribuição de probabilidade conjunta, consideramos θ como um valor fixo e os dados como variáveis aleatórias.
- Isso representa a probabilidade de observar um conjunto específico de valores dos dados dado o parâmetro θ.

Como Aprender sobre um Parâmetro θ?

1. Traduza seu conhecimento *a priori* sobre θ em uma distribuição de probabilidade sobre θ, $p(\theta)$.
2. Colete dados e forme a função de verossimilhança $p(\text{dados}|\theta)$.
3. Uma posteriori pode ser usada como priori em novas análises, com novas evidências. ### Conclusão

O teorema de Bayes permite combinar conhecimento *a priori* com dados observados para atualizar a crença sobre um parâmetro ou evento. É uma ferramenta fundamental em estatística bayesiana para inferência e tomada de decisões.

Bayesian Process Tracing

- Especifique hipóteses H_i e suas prioris $P(H_i)$.
- Identifique as evidências disponíveis e construa uma verossimilhança $P(E|H_i)$ para cada hipótese i.
- Obtenha *posterior odds* em comparações pareadas de hipóteses: $\frac{P(H_i|E)}{P(H_j|E)} = \frac{P(H_i)P(E|H_i)}{P(H_j)P(E|H_j)}$

Componentes críticos

- Como definir as prioris
- O que são evidências
- Como construir as verossimilhanças
- Hipóteses precisam ser rivais, isto é, $P(H_i) + P(H_j) = 1$.

Prioris

- Longa histórica sobre definição de prioris
- Elicitação de experts
- Usar prioris "não-informativas"
- Análise de sensibilidade (outras prioris mudariam a conclusão)?

Evidências

- Não é trivial, em process Tracing Bayesiano ou não, definir o que são evidências distintas. Uma informação de uma fonte e matéria de jornal com a mesma fonte são duas evidências distintas? Duas observações (ainda que correlacionadas) ou a mesma evidência?
- Tradicionalmente, apenas evidência within-case (que seria o objeto de Process Tracing). No Bayesianismo, pode ser evidências de outros casos similares, por exemplo.
- Heurística: evidências que favorecem hipóteses distintas ou de diferentes tipos de fontes (ex. Bolsonarista e petista) devem ser consideradas evidências distintas. Já informações similares de fontes similares (exemplos, dois membros do governo contam a mesma história) devem ser consideradas a mesma evidência.

E não desagregue a evidência demais, para não dificultar a quantificação da verossimilhança.

Verossimilhanças

- Intuição: Em um mundo em que H_i é verdade, quanto estaríamos surpresos ou seria esperado observar E?
- Uso de logaritmos (decibeis) são úteis para calibrar o peso das evidências na razão de verosimilhanças: $10 \cdot log_{10}(P(E|H_i)/P(E|H_j))$. Chamado de peso da evidência
- Ideia é aproximar os decibéis (que são calculados em escala logaritmas). Fairfield & Charman (2017) (2017) recomendam que o mínimo distinguível é $1db$.
- Bennet considera um smoking gun se $P(E|H_j) = .05$ e $P(E|H_i) = .2$. Isso dá $6db$. Segundo Charmman & Fairfiel (2017), saliente, mas longe de um *smokinggun*.

Hipóteses

- Hipóteses precisam ser rivais, isto é, uma não pode conter a outra.
- A mera negação lógica não é recomendado, pois inclui infinitas hipóteses, várias contraditórias entre si.
- Múltiplas causas tornam difícil construir hipóteses rivais

Limites

- Fairfield & Charman (2017) falam que workshops de 1 ou 2 dias não são suficientes para treinar pesquisadores.
- Em muitos casos é difícil especificar probabilidades precisamente e podemos passar impressão de precisão onde ela não existe.
- Às vezes hipóteses explicativas envolvem interação ou complixidades que tornam difícil operacionalizar hipóteses rivais.
- Não deve substituir narrativa de casos e pode se tornar muito demandante aplicar process tracing Bayesiano para todos os casos analisados em um dado contexto.