Introdução à Amostragem para Inferência Bayesiana - Aula 2

Eliezer de Souza da Silva, PhD
EMAp FGV / BCAM / UFC
eliezer@probabilistic.ai
https:\sereliezer.github.io

4 de Fevereiro de 2025

Plano

- **Aula anterior**: modelos conjugados e não-conjugados, métodos de amostragem por inversão da cumulativa e amostragem por rejeição.
- **Hoje**:
  - O que é uma cadeia de Markov e porque estamos estudando esse objeto matemático?
  - **Algoritmos de MCMC**: Metropolis-Hasting, Langevin MCMC, Gibbs Sampling, Hamiltonian Monte Carlo.
  - Pontos de interseção: descida de gradiente e hamiltonian monte carlo; treinamento de redes neurais bayesianas.
  - Considerações em espaços multidimensionais contínuos, discretos e híbridos.
  - **Extras**:
    - Espaços sequenciais: filtros de partículas / Sequential Monte Carlo.
    - Pontos de interseção: busca e amostragem em espaços discretos composicionais e complexos (árvores, redes).
    - Espaços com geometria não-euclidiana.

O que é Inferência Bayesiana?

Modelagem: Conjunto de dados observados $D \subset \mathcal{X}$ para algum domínio $\mathcal{X}$.

- $D = \{X_i, \ldots, X_n\}$ ou $D = \{(X_i, Y_i), \ldots, (X_n, Y_n)\}$.

O que é Inferência Bayesiana?

Modelagem: Conjunto de dados observados $D \subset \mathcal{X}$ para algum domínio $\mathcal{X}$.

- $D = \{X_i, \ldots, X_n\}$ ou $D = \{(X_i, Y_i), \ldots, (X_n, Y_n)\}$.
- $\theta \in \Theta$ para algum domínio de parâmetros.

O que é Inferência Bayesiana?

Modelagem: Conjunto de dados observados $D \subset \mathcal{X}$ para algum domínio $\mathcal{X}$.

- $D = \{X_i, \ldots, X_n\}$ ou $D = \{(X_i, Y_i), \ldots, (X_n, Y_n)\}$.
- $\theta \in \Theta$ para algum domínio de parâmetros.

Inferência Bayesiana: Uso da regra de Bayes para atualizar crenças com base em dados observados.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{1}$$

O que é Inferência Bayesiana?

Modelagem: Conjunto de dados observados $D \subset \mathcal{X}$ para algum domínio $\mathcal{X}$.

- $D = \{X_i, \ldots, X_n\}$ ou $D = \{(X_i, Y_i), \ldots, (X_n, Y_n)\}$.
- $\theta \in \Theta$ para algum domínio de parâmetros.

Inferência Bayesiana: Uso da regra de Bayes para atualizar crenças com base em dados observados.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{1}$$

Problema: O cálculo da integral $P(D) = \int P(D|\theta)P(\theta)d\theta$ pode ser intratável. Exemplo: amostrar de uma distribuição não normalizada

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}}\right) \times \left(\lambda^{Y-1}e^{-n\lambda}\right)$$

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas, especialmente quando a normalização é intratável.

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas, especialmente quando a normalização é intratável.

Principais Ideias:

- Utiliza cadeias de Markov para gerar amostras dependentes.
- Converge para a distribuição alvo após um número suficiente de iterações.

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas, especialmente quando a normalização é intratável.

Principais Ideias:

- Utiliza cadeias de Markov para gerar amostras dependentes.
- Converge para a distribuição alvo após um número suficiente de iterações.

Exemplos de Algoritmos MCMC:

- Metropolis-Hastings: propõe novos pontos e aceita com uma certa taxa.
- Gibbs Sampling: amostragem condicional de uma variável por vez.
- Hamiltonian Monte Carlo (HMC): usa gradientes para explorar o espaço de amostragem de maneira mais eficiente.

Cadeias de Markov

Definição: Um processo estocástico onde o próximo estado depende apenas do estado atual.

Cadeias de Markov

Definição: Um processo estocástico onde o próximo estado depende apenas do estado atual.

Propriedades:

- Matriz ou kernel de transição P. Com $P(X_j, X_i)$ representando a probabilidade de transição do estado X_i para X_j

Cadeias de Markov

Definição: Um processo estocástico onde o próximo estado depende apenas do estado atual.
Propriedades:

- Matriz ou kernel de transição P. Com $P(X_j, X_i)$ representando a probabilidade de transição do estado X_i para X_j
- Estado estacionário: $\pi P = \pi$.

Cadeias de Markov

Definição: Um processo estocástico onde o próximo estado depende apenas do estado atual.

Propriedades:

- Matriz ou kernel de transição P. Com $P(X_j, X_i)$ representando a probabilidade de transição do estado X_i para X_j
- Estado estacionário: $\pi P = \pi$.
- Propriedade de equilíbrio detalhado: Se uma cadeia de Markov satisfaz a propriedade de equilíbrio detalhado, então para quaisquer estados $X, X' \in \mathcal{X}$, temos:

$$\pi(X)P(X, X') = \pi(X')P(X', X) \tag{2}$$

Cadeias de Markov

Definição: Um processo estocástico onde o próximo estado depende apenas do estado atual.

Propriedades:

- Matriz ou kernel de transição P. Com $P(X_j, X_i)$ representando a probabilidade de transição do estado X_i para X_j
- Estado estacionário: $\pi P = \pi$.
- Propriedade de equilíbrio detalhado: Se uma cadeia de Markov satisfaz a propriedade de equilíbrio detalhado, então para quaisquer estados $X, X' \in \mathcal{X}$, temos:

$$\pi(X)P(X, X') = \pi(X')P(X', X) \tag{2}$$

Essa propriedade implica que a distribuição π permanece estável ao longo de iterações da matrix de transição, garantindo a convergência para a distribuição estacionária.

Reversibilidade em Cadeias de Markov

Definição: Seja P um kernel de Markov em um espaço $\mathcal{X} \times \mathcal{X}$. Uma medida ξ sobre $\mathcal{X}$ é dita reversível com respeito a P se a medida produto $\xi \otimes P$ sobre $\mathcal{X} \times \mathcal{X}$ for simétrica, ou seja, para quaisquer conjuntos mensuráveis $A, B \subset \mathcal{X}$:

$$\xi(A) \otimes P(A \times B) = \xi(B) \otimes P(B \times A). \tag{3}$$

Isso significa que a probabilidade de transição de X para X' sob ξ e P é a mesma de X' para X,, garantindo equilíbrio estatístico entre estados.

Condição de Equilíbrio Detalhado

Definição: A reversibilidade implica a condição de equilíbrio detalhado, que exige que para todos os estados $(X, X') \in \mathcal{X} \times \mathcal{X}$:

$$\xi(X)P(X, X') = \xi(X')P(X', X). \tag{4}$$

Essa condição assegura que o fluxo de probabilidade entre quaisquer dois estados é equilibrado, garantindo que a distribuição de estados permaneça inalterada.

Condição de Equilíbrio Detalhado

Definição: A reversibilidade implica a condição de equilíbrio detalhado, que exige que para todos os estados $(X, X') \in \mathcal{X} \times \mathcal{X}$:

$$\xi(X)P(X, X') = \xi(X')P(X', X). \tag{4}$$

Essa condição assegura que o fluxo de probabilidade entre quaisquer dois estados é equilibrado, garantindo que a distribuição de estados permaneça inalterada.

Implicação na Inferência Bayesiana:

- Em algoritmos MCMC, a reversibilidade assegura que as amostras geradas convergem corretamente para a distribuição de interesse.
- A propriedade de equilíbrio detalhado evita vieses sistemáticos na exploração do espaço amostral.

Amostragem a partir da Distribuição Estacionária

Conclusão:

- A condição de equilíbrio detalhado implica que a distribuição estacionária π é uma solução do sistema de equações lineares $\pi P = \pi$.
- Isso significa que, ao longo de iterações suficientes, um processo MCMC produzirá amostras da distribuição alvo.
- Métodos como Metropolis-Hastings e Gibbs Sampling garantem a reversibilidade e convergência para a distribuição estacionária. Alternativamente, podemos projetar kernel de transição que respeite essa condição, com esquemas de *amostragem por rejeição*.

Amostragem de Gibbs

Definição: A amostragem de Gibbs é um método MCMC que amostra sequencialmente de distribuições condicionais completas.

Amostragem de Gibbs

Definição: A amostragem de Gibbs é um método MCMC que amostra sequencialmente de distribuições condicionais completas.

Passos:

1. Inicializar $X^{(0)}$.
2. Para cada iteração t, amostrar sequencialmente de cada variável condicional:
$$X_i^{(t+1)} \sim p(X_i | X_{-i}^{(t)}), \tag{5}$$
onde X_{-i} representa todas as variáveis exceto X_i.

Amostragem de Gibbs

Preservação do Equilíbrio Detalhado: O equilíbrio detalhado é garantido porque:

- Cada variável é amostrada condicionalmente, garantindo que a distribuição conjunta seja preservada.
- A cadeia de Markov gerada por Gibbs Sampling é reversível, pois a probabilidade de transição entre estados satisfaz:

$$p(X'|X)p(X) = p(X|X')p(X'). \tag{6}$$

- Essa reversibilidade assegura a convergência para a distribuição estacionária desejada.

Amostragem de Gibbs

Preservação do Equilíbrio Detalhado: O equilíbrio detalhado é garantido porque:

- Cada variável é amostrada condicionalmente, garantindo que a distribuição conjunta seja preservada.
- A cadeia de Markov gerada por Gibbs Sampling é reversível, pois a probabilidade de transição entre estados satisfaz:

$$p(X'|X)p(X) = p(X|X')p(X'). \tag{6}$$

- Essa reversibilidade assegura a convergência para a distribuição estacionária desejada.

Aplicações:

- Modelos Bayesianos Hierárquicos.
- Inferência em modelos gráficos probabilísticos.
- Problemas de otimização combinatória.

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas.

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas.

Principais Ideias:

- Utiliza cadeias de Markov para gerar amostras dependentes.
- Converge para a distribuição alvo após várias iterações.

Introdução ao MCMC

Markov Chain Monte Carlo (MCMC): Uma classe de métodos para amostragem de distribuições complexas.

Principais Ideias:

- Utiliza cadeias de Markov para gerar amostras dependentes.
- Converge para a distribuição alvo após várias iterações.

Algoritmos MCMC:

- **Metropolis-Hastings**: aceita ou rejeita novas amostras com certa taxa

Algoritmo Metropolis-Hastings

Desenvolvimento: O algoritmo Metropolis-Hastings é construído a partir da propriedade de reversibilidade, garantindo que a distribuição estacionária $\pi(x)$ seja preservada ao longo da cadeia de Markov.

Algoritmo Metropolis-Hastings

Desenvolvimento: O algoritmo Metropolis-Hastings é construído a partir da propriedade de reversibilidade, garantindo que a distribuição estacionária $\pi(x)$ seja preservada ao longo da cadeia de Markov.

Passos:

1. Propor um novo estado $x' \sim q(x'|x)$.
2. Calcular a taxa de aceitação:
$$\alpha = \min\left(1, \frac{p(x')q(x|x')}{p(x)q(x'|x)}\right). \tag{7}$$
3. Aceitar x' com probabilidade α, senão manter x.

Algoritmo Metropolis-Hastings

Propriedades da Proposta $q(x'|x)$:

- Deve garantir ergodicidade para que todo espaço amostral seja explorado.
- Pode ser simétrica $(q(x'|x) = q(x|x'))$ ou assimétrica, desde que a taxa de aceitação corrija o viés.
- Distribuições propostas eficientes devem equilibrar exploração e aceitação.

Taxa de Aceitação e Equilíbrio Detalhado

Relação com Reversibilidade: A taxa de aceitação α é derivada impondo a condição de equilíbrio detalhado:

$$\pi(x)P(x \to x') = \pi(x')P(x' \to x), \tag{8}$$

substituindo $P(x \to x') = q(x'|x)\alpha(x, x')$:

$$\pi(x)q(x'|x)\alpha(x, x') = \pi(x')q(x|x')\alpha(x', x). \tag{9}$$

Isso leva à escolha:

$$\alpha(x, x') = \min\left(1, \frac{p(x')q(x|x')}{p(x)q(x'|x)}\right), \tag{10}$$

assegurando a reversibilidade e convergência para a distribuição estacionária.

Exemplo: Distribuição Alvo

Consideramos a distribuição alvo:

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log \lambda - \mu)^2}{2\sigma^2}} \right) \times \left(\lambda^{Y-1} e^{-n\lambda} \right). \tag{11}$$

Exemplo: Distribuição Alvo

Consideramos a distribuição alvo:

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}}\right) \times \left(\lambda^{Y-1} e^{-n\lambda}\right). \tag{11}$$

Procedimento: https://colab.research.google.com/drive/1hFzEL-rOeetqqJkNKvSh1QhqUBGxtSUcw?usp=sharing

- Escolhemos uma distribuição proposta $q(\lambda'|\lambda) = \mathcal{N}(\log\lambda'; \log\lambda, \tau^2)$.
- Calculamos a taxa de aceitação α usando a fórmula do Metropolis-Hastings.
- Iteramos para obter amostras da distribuição alvo não normalizada.

Amostragem de Distribuições Não Normalizadas

Motivação: Em muitos problemas de inferência Bayesiana, a distribuição posterior é conhecida apenas até uma constante de normalização inatingível.

Amostragem de Distribuições Não Normalizadas

Motivação: Em muitos problemas de inferência Bayesiana, a distribuição posterior é conhecida apenas até uma constante de normalização inatingível.

Solução com Metropolis-Hastings:

- O algoritmo requer apenas a razão das probabilidades $p(x')/p(x)$, eliminando a necessidade de normalização.
- Essa propriedade permite amostragem eficiente de distribuições complexas, como na inferência Bayesiana com priors não conjugados.
- A aceitação seletiva baseada na reversibilidade garante que a cadeia de Markov ainda converge para a distribuição correta.

Langevin Dynamics - Introdução

Motivação: Queremos gerar amostras de uma distribuição alvo $\pi(x)$ através de um processo baseado em difusão estocástica.

Langevin Dynamics - Introdução

Motivação: Queremos gerar amostras de uma distribuição alvo $\pi(x)$ através de um processo baseado em difusão estocástica.
Definição: O processo de Langevin é definido pela equação diferencial estocástica:

$$\dot{X} = \nabla \log \pi(X) + \sqrt{2}\dot{W}, \tag{12}$$

onde $\dot{W}$ é um processo de Wiener (movimento Browniano padrão).

Discretização via Euler–Maruyama

Aproximação Numérica: Podemos simular esse processo usando o método de Euler–Maruyama com um passo τ:

$$X_{k+1} = X_k + \tau \nabla \log \pi(X_k) + \sqrt{2\tau}\xi_k, \tag{13}$$

com $\xi_k \sim \mathcal{N}(0, I)$. Esse método permite gerar aproximações da dinâmica contínua de Langevin.

Estacionaridade e Amostragem Correta

Propriedade Fundamental: No limite $t \to \infty$, a distribuição ρ_t da variável $X(t)$ converge para a distribuição estacionária desejada:

$$\rho_\infty = \pi. \tag{14}$$

Isso garante que podemos utilizar a dinâmica de Langevin para amostrar corretamente de distribuições alvo complexas.

Exemplo Computacional - Langevin Dynamics

Implementação: Utilizamos a seguinte discretização para gerar amostras da distribuição alvo:

- Cálculo do gradiente $\nabla \log \pi(X_k)$.
- Atualização da posição baseada na discretização de Euler–Maruyama.
- Introdução de ruído gaussiano para garantir a difusão correta.

Comparação entre Langevin Dynamics e SGD

Semelhanças:

- Ambos utilizam gradientes para guiar a atualização dos parâmetros.
- Ambos podem ser interpretados como processos de otimização.

Comparação entre Langevin Dynamics e SGD

Semelhanças:

- Ambos utilizam gradientes para guiar a atualização dos parâmetros.
- Ambos podem ser interpretados como processos de otimização.

Diferenças:

- O SGD (Stochastic Gradient Descent) minimiza uma função de perda determinística, enquanto Langevin Dynamics incorpora ruído estocástico para garantir amostragem correta.
- Langevin Dynamics converge para a distribuição alvo $\pi(x)$, enquanto SGD busca um ponto ótimo único.
- O ruído adicional em Langevin Dynamics melhora a exploração do espaço de estados, prevenindo estagnação em mínimos locais.

Metropolis-Adjusted Langevin Algorithm (MALA)

Extensão de Langevin Dynamics: O MALA combina a atualização de Langevin com um critério de aceitação-rejeição do Metropolis-Hastings.

Metropolis-Adjusted Langevin Algorithm (MALA)

Extensão de Langevin Dynamics: O MALA combina a atualização de Langevin com um critério de aceitação-rejeição do Metropolis-Hastings.
Discretização:

$$X_{k+1} = X_k + \frac{\tau}{2} \nabla \log \pi(X_k) + \sqrt{\tau} \xi_k, \tag{15}$$

onde $\xi_k \sim \mathcal{N}(0, I)$. A proposta é aceita com probabilidade:

$$\alpha = \min \left(1, \frac{\pi(X_{k+1}) q(X_k | X_{k+1})}{\pi(X_k) q(X_{k+1} | X_k)} \right). \tag{16}$$

Metropolis-Adjusted Langevin Algorithm (MALA)

Extensão de Langevin Dynamics: O MALA combina a atualização de Langevin com um critério de aceitação-rejeição do Metropolis-Hastings.
Discretização:

$$X_{k+1} = X_k + \frac{\tau}{2}\nabla \log \pi(X_k) + \sqrt{\tau}\xi_k, \tag{15}$$

onde $\xi_k \sim \mathcal{N}(0, I)$. A proposta é aceita com probabilidade:

$$\alpha = \min\left(1, \frac{\pi(X_{k+1})q(X_k|X_{k+1})}{\pi(X_k)q(X_{k+1}|X_k)}\right). \tag{16}$$

Benefícios:

- Melhor exploração do espaço de estados do que Langevin Dynamics puro.
- Reduz viés introduzido pela discretização.
- Útil para distribuições com caudas longas ou formas complexas.

Mecânica Hamiltoniana:

- Baseia-se na função Hamiltoniana, $H(q, p) = T(p) + V(q)$, onde T é a energia cinética e V é a energia potencial.
- As equações de Hamilton governam a evolução temporal do sistema:

$$\dot{q} = \frac{\partial H}{\partial p}, \tag{17}$$

$$\dot{p} = -\frac{\partial H}{\partial q}. \tag{18}$$

- Essas equações permitem explorar eficientemente o espaço de amostragem.

Mecânica Hamiltoniana e Princípio da Ação Mínima II

Princípio da Ação Mínima:

- O movimento de um sistema segue um caminho que minimiza a ação S, definida como a integral do lagrangiano ao longo do tempo:

$$S = \int L(q, \dot{q}) dt. \tag{19}$$

- Em HMC, isso garante que os passos seguem trajetórias naturais, reduzindo rejeições.

Hamiltonian Monte Carlo (HMC)

Ideia Principal: Usa conceitos da mecânica hamiltoniana para explorar o espaço de amostragem.

Hamiltonian Monte Carlo (HMC)

Ideia Principal: Usa conceitos da mecânica hamiltoniana para explorar o espaço de amostragem.

Passos:

- Introduz variáveis auxiliares de momento p.
- Define energia total: $H(x, p) = U(x) + K(p)$.
- Evolui usando equações de Hamilton.
- Aceita/rejeita novo estado.

Hamiltonian Monte Carlo (HMC)

Ideia Principal: Usa conceitos da mecânica hamiltoniana para explorar o espaço de amostragem.

Passos:

- Introduz variáveis auxiliares de momento p.
- Define energia total: $H(x, p) = U(x) + K(p)$.
- Evolui usando equações de Hamilton.
- Aceita/rejeita novo estado.

Benefícios:

- Move-se eficientemente em espaços de alta dimensão.
- Evita passos pequenos do Metropolis-Hastings.

Cálculo da Taxa de Aceitação no HMC

Critério de Aceitação:

- Em HMC, um novo estado (q', p') é aceito com a seguinte taxa:

$$\alpha = \min\left(1, \frac{\exp(-H(q', p'))}{\exp(-H(q, p))}\right). \tag{20}$$

- Como $H(q, p)$ é aproximadamente conservado durante as simulações, as taxas de aceitação são tipicamente altas.

Cálculo da Taxa de Aceitação no HMC

Critério de Aceitação:

- Em HMC, um novo estado (q', p') é aceito com a seguinte taxa:

$$\alpha = \min\left(1, \frac{\exp(-H(q', p'))}{\exp(-H(q, p))}\right). \tag{20}$$

- Como $H(q, p)$ é aproximadamente conservado durante as simulações, as taxas de aceitação são tipicamente altas.

Relação com Metropolis-Hastings:

- No Metropolis-Hastings tradicional, a taxa de aceitação depende da razão de densidades da posterior.
- Em HMC, a proposta de novos estados é guiada por equações diferenciais, reduzindo rejeições e tornando a amostragem mais eficiente.

Descida de Gradiente e Hamiltonian Monte Carlo

Interseção: Ambos utilizam gradientes para guiar a exploração do espaço de estados.

Descida de Gradiente e Hamiltonian Monte Carlo

Interseção: Ambos utilizam gradientes para guiar a exploração do espaço de estados.

Diferenças:

- Descida de Gradiente: usada para otimização, busca um mínimo da função objetivo.
- Hamiltonian Monte Carlo (HMC): amostragem eficiente, utilizando energia potencial e cinética.
- HMC evita passos pequenos da descida de gradiente ao introduzir variáveis auxiliares de momento.

Treinamento de Redes Neurais Bayesianas

Abordagem: Modelamos pesos das redes neurais como distribuições probabilísticas.

Treinamento de Redes Neurais Bayesianas

Abordagem: Modelamos pesos das redes neurais como distribuições probabilísticas.

Métodos Populares:

- Amostragem MCMC para pesos da rede.
- Variational Inference para aproximações eficientes.
- Hamiltonian Monte Carlo para aprendizado profundo Bayesiano.

Considerações sobre Espaços Multidimensionais

Tipos de Espaços:

- Contínuos: distribuição Gaussiana, HMC para amostragem eficiente.
- Discretos: modelos gráficos probabilísticos, Gibbs Sampling.
- Híbridos: combinação de espaços contínuos e discretos, aplicações em inferência Bayesiana complexa.

Filtros de Partículas e Sequential Monte Carlo

Motivação: Métodos de inferência para modelos dinâmicos latentes.

Motivação: Métodos de inferência para modelos dinâmicos latentes.
Passos:

- Inicializar partículas representando estados ocultos.
- Atualizar pesos das partículas conforme novas observações chegam.
- Reamostrar partículas para manter diversidade.

Busca e Amostragem em Espaços Discretos Composicionais

Exemplos:

- Amostragem em árvores bayesianas.
- Inferência probabilística em redes estruturadas.
- Modelos de grafos probabilísticos.

Busca e Amostragem em Espaços Discretos Composicionais

Exemplos:

- Amostragem em árvores bayesianas.
- Inferência probabilística em redes estruturadas.
- Modelos de grafos probabilísticos.

Métodos:

- Monte Carlo Tree Search (MCTS).
- Métodos MCMC adaptativos para grafos.

Amostragem em Espaços Não-Euclideanos

Motivação: Muitos problemas têm estrutura geométrica não trivial.

Amostragem em Espaços Não-Euclideanos

Motivação: Muitos problemas têm estrutura geométrica não trivial.
Exemplos:

- Inferência Bayesiana em variedades Riemannianas.
- Hamiltonian Monte Carlo em espaços curvos.
- Aprendizado de representação em espaços hiperbólicos.

Amostragem em Espaços Não-Euclideanos

Equações do HMC em Espaços Curvos: A evolução das dinâmicas de Hamilton é governada pelas equações:

$$\frac{dq^i}{dt} = g^{ij} \frac{\partial H}{\partial p^j}, \tag{21}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i} + \frac{1}{2} \frac{\partial g^{jk}}{\partial q^i} p_j p_k, \tag{22}$$

onde g^{ij} é a métrica da variedade e H é a Hamiltoniana do sistema.

Conclusões e Discussão

Resumo dos Métodos:

- Metropolis-Hastings e HMC permitem amostrar distribuições complexas.
- Langevin Dynamics e MALA combinam gradientes e difusão para melhor eficiência.
- Gibbs Sampling é útil para distribuições condicionais conhecidas.

Conclusões e Discussão

Resumo dos Métodos:

- Metropolis-Hastings e HMC permitem amostrar distribuições complexas.
- Langevin Dynamics e MALA combinam gradientes e difusão para melhor eficiência.
- Gibbs Sampling é útil para distribuições condicionais conhecidas.

Linguagens Probabilísticas e Bibliotecas:

- Probabilistic Programming Languages (PPLs) como Pyro, Stan e Turing.jl oferecem abstrações poderosas.
- Essas ferramentas permitem especificar modelos complexos e realizar inferência automaticamente.
- Integração com métodos de amostragem avançados facilita experimentação e análise.