Aprendizagem Espectral em Modelos Ocultos de Markov

Experimentos com a nova abordagem

Jonas Rocha Lima Amaro

Modelos Ocultos de Markov

•

Aprendizagem Espectral

•

Experimentos

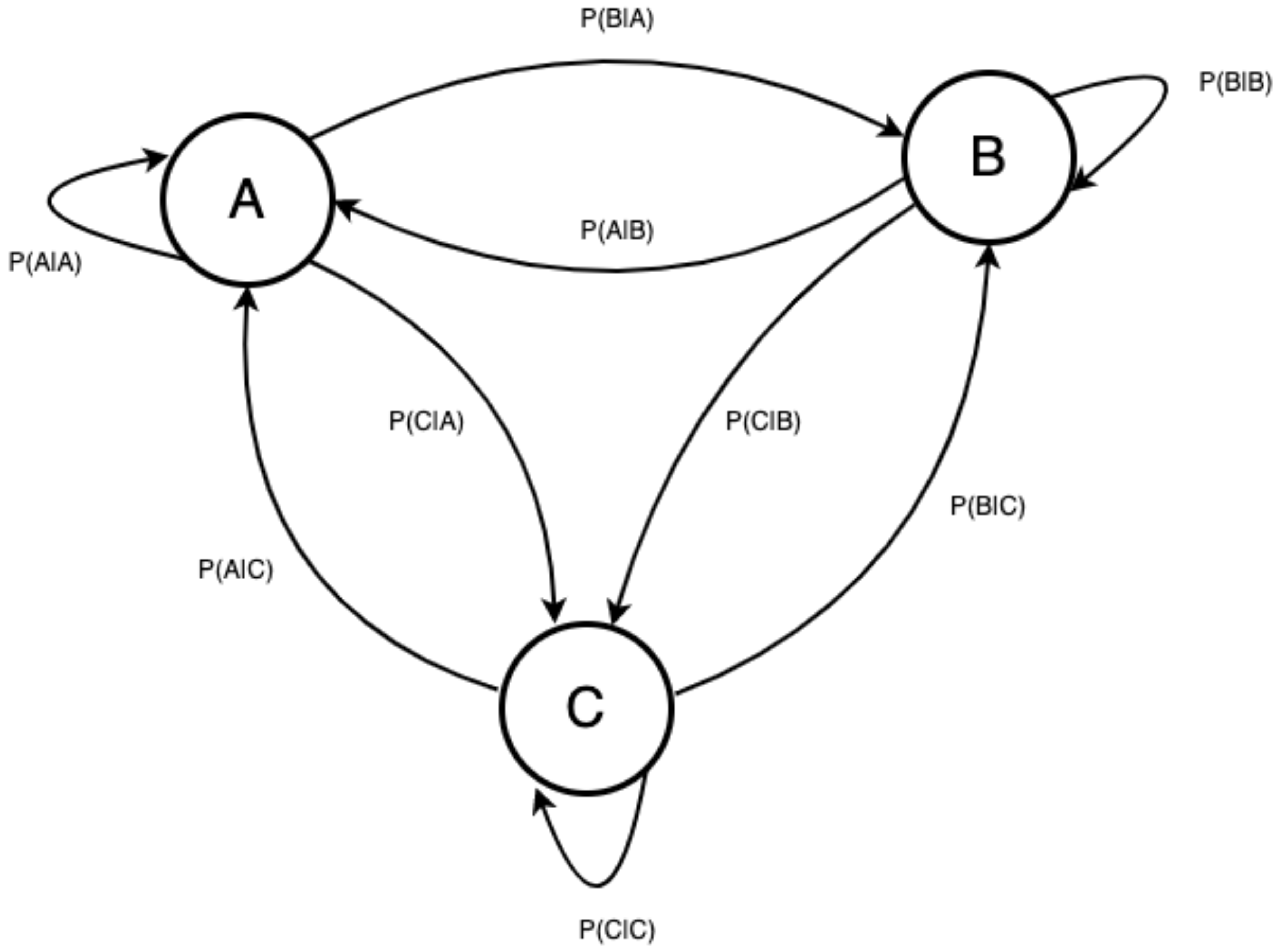
Modelos Ocultos de Markov

- Exemplo inicial
- Parâmetros do Modelo
- Aplicações do Modelo

Exemplo Inicial

Urnas e Bolas

- Sorteio de sequências de bolas, na moda markoviana
- Processo com estados ocultos e símbolos discretos
- Quando dá para ter certeza de que urna veio cada bola?



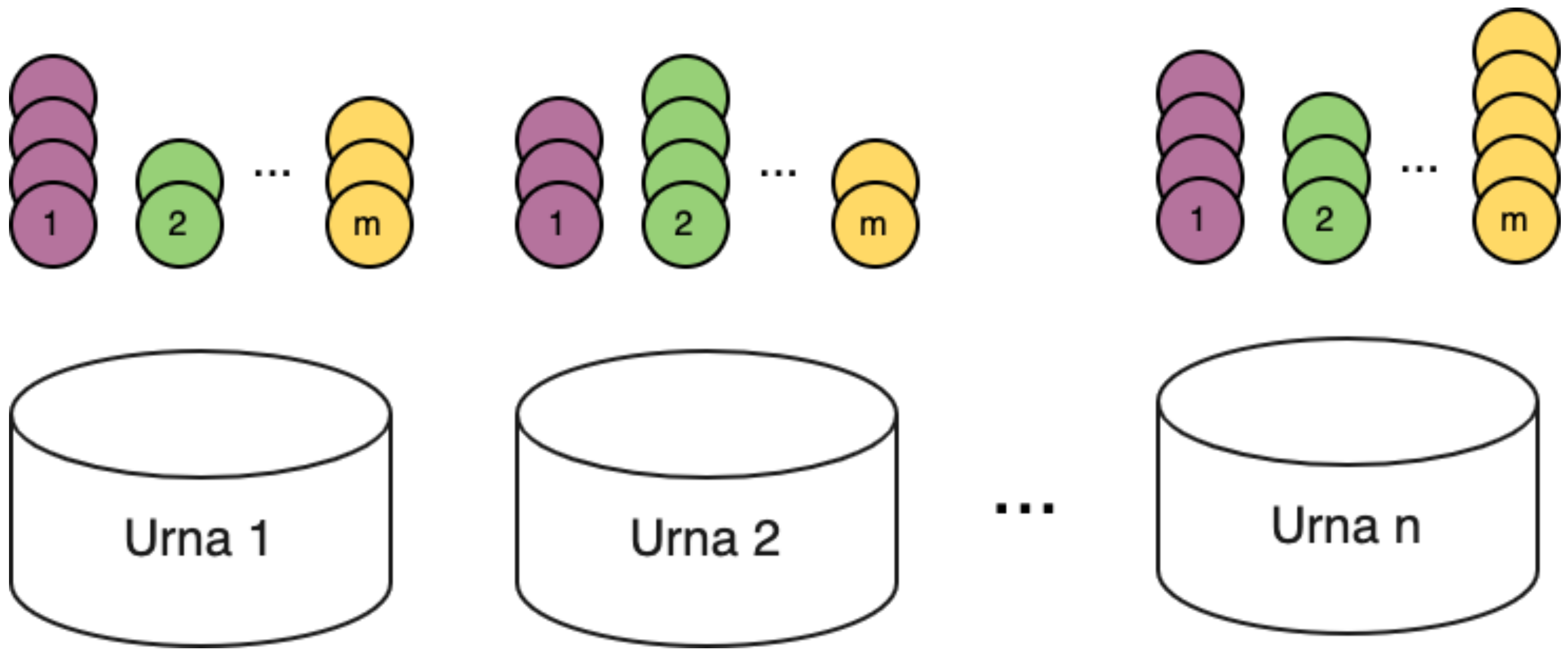
Exemplo Inicial

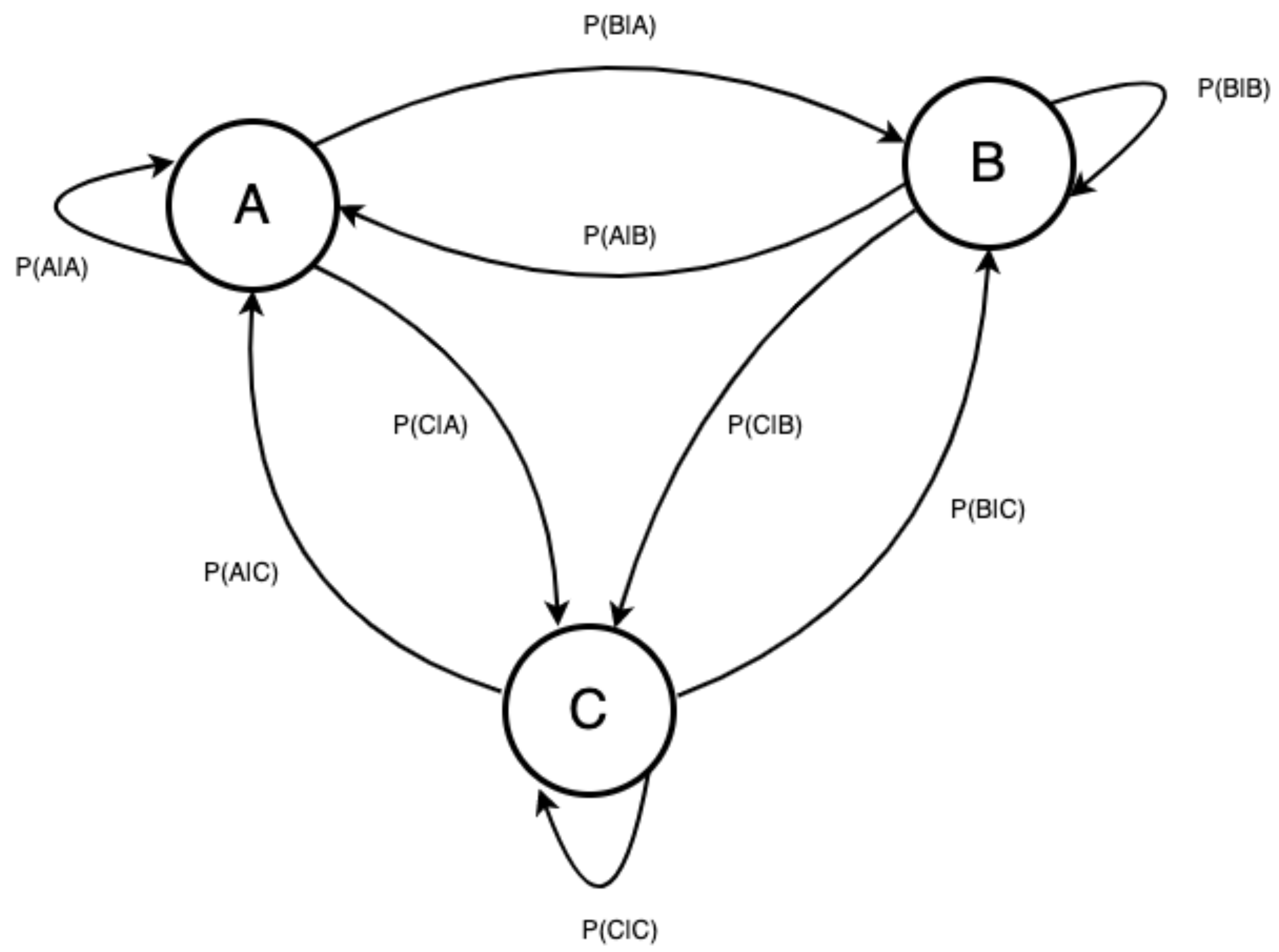
Urnas e Bolas

- Sorteio de sequências de bolas, na moda markoviana

- Processo com estados ocultos e símbolos discretos

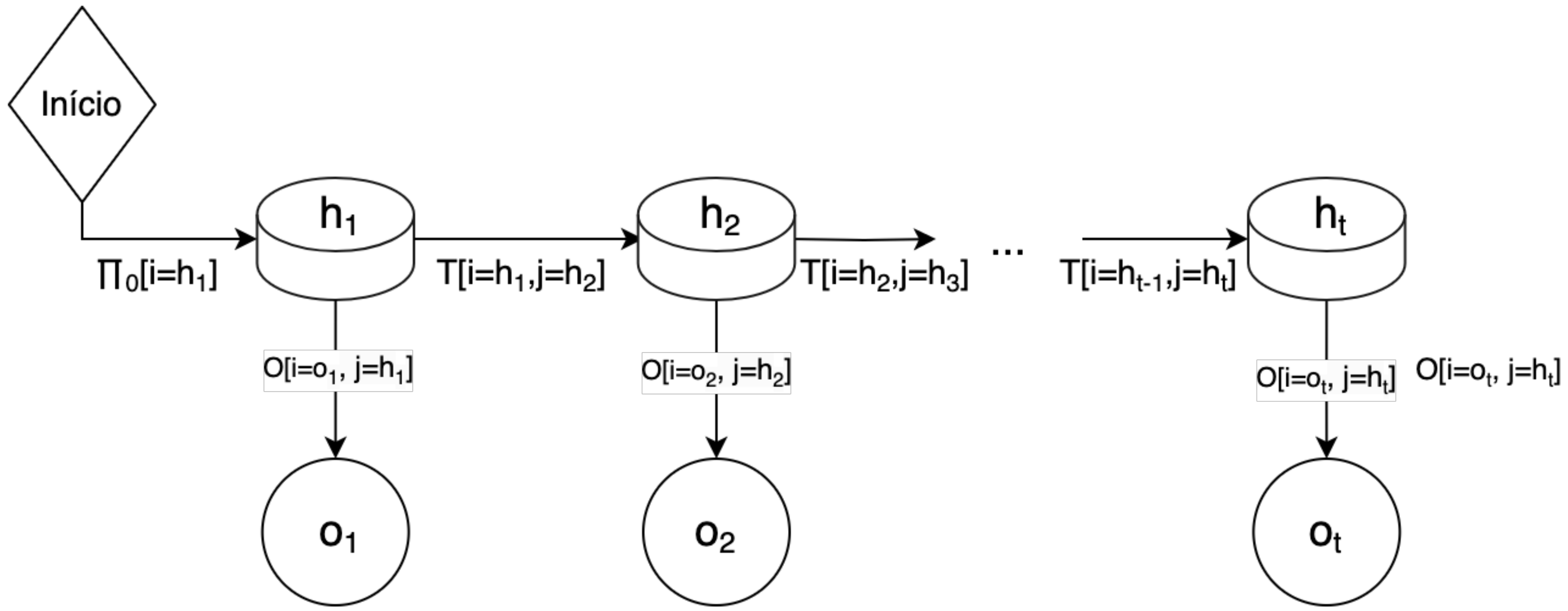
- Quando dá para ter certeza de que urna veio cada bola?





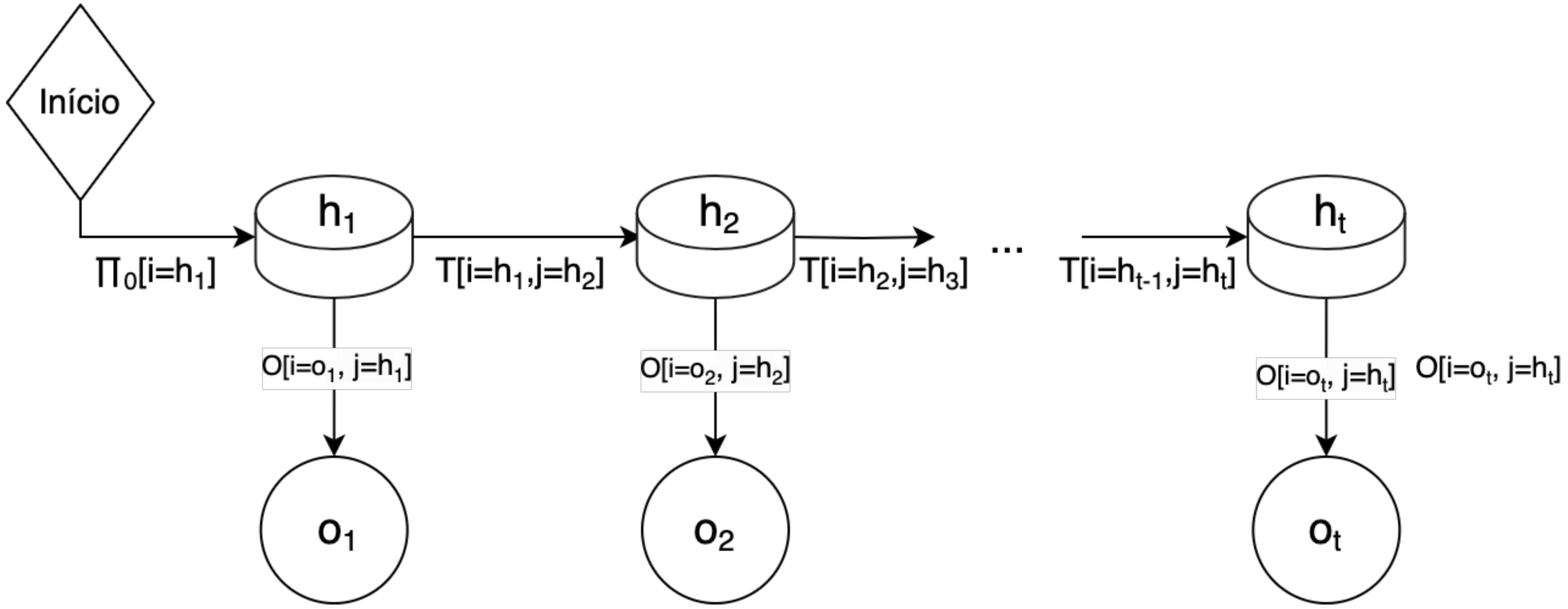
Parâmetros do Modelo

- $n \in \mathbb{N}$ Número de estados ocultos



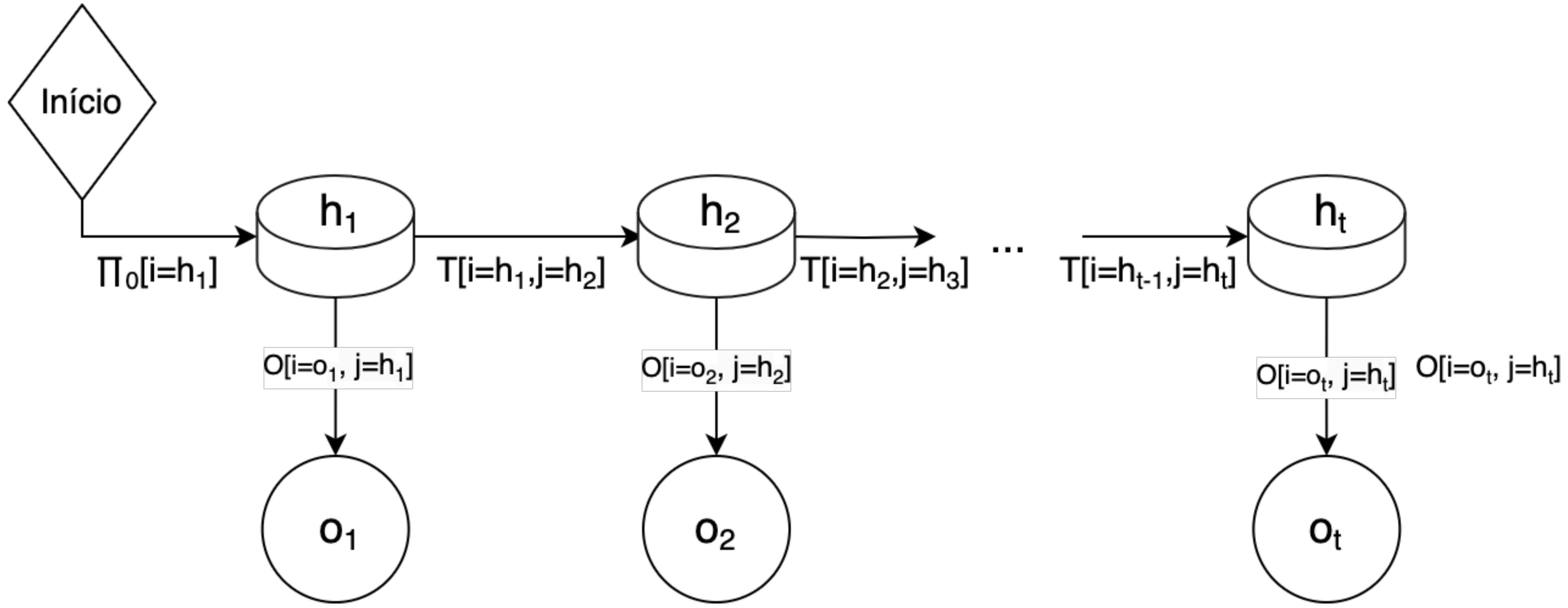
Parâmetros do Modelo

- $n \in \mathbb{N}$ Número de estados ocultos
- $m \in \mathbb{N}$ Número de símbolos observáveis



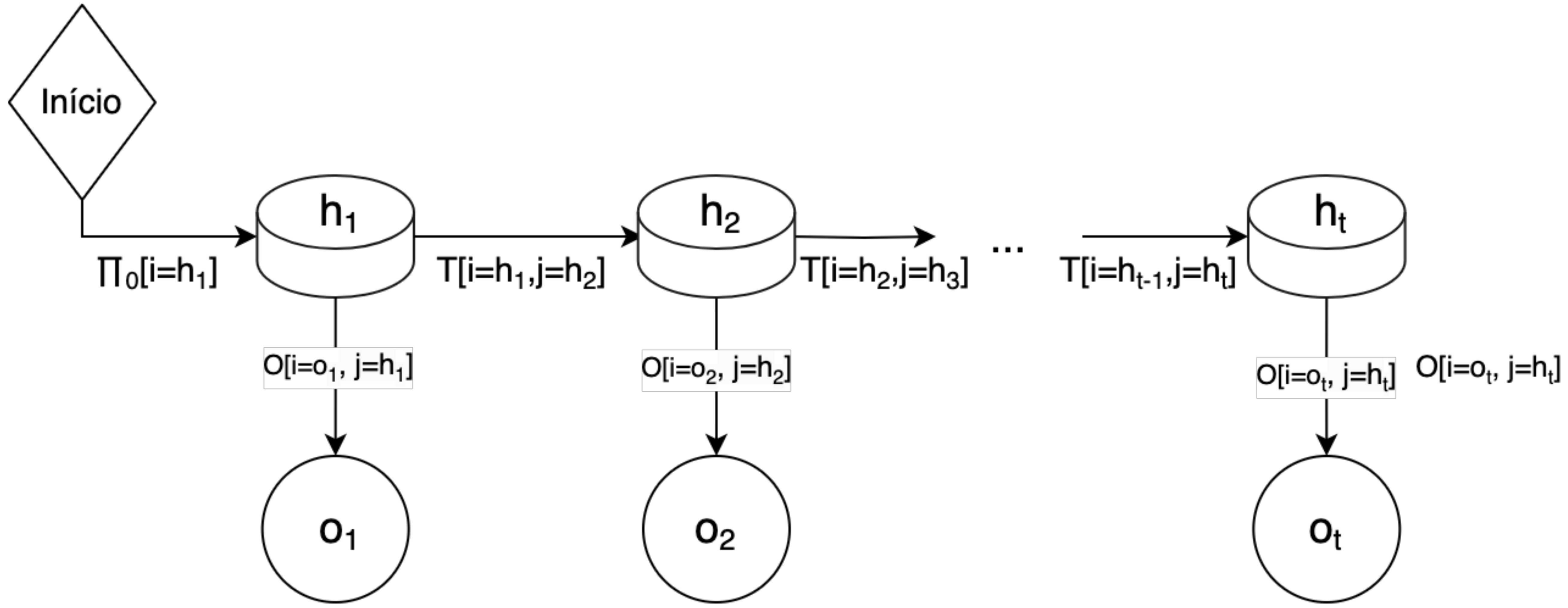
Parâmetros do Modelo

- $n \in \mathbb{N}$ Número de estados ocultos
- $m \in \mathbb{N}$ Número de símbolos observáveis
- $\Pi_0 \in \mathbb{R}^n$ Distribuição do estado inicial



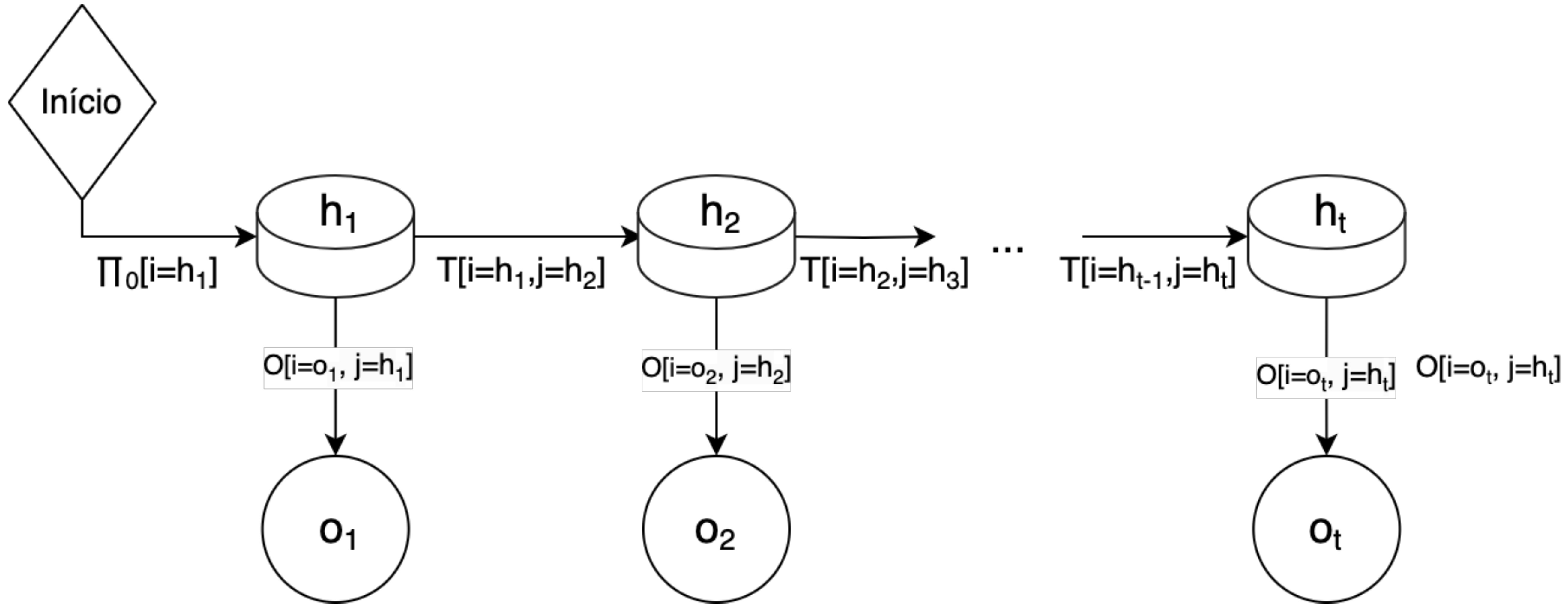
Parâmetros do Modelo

- $n \in \mathbb{N}$ Número de estados ocultos
- $m \in \mathbb{N}$ Número de símbolos observáveis
- $\Pi_0 \in \mathbb{R}^n$ Distribuição do estado inicial
- $O \in \mathbb{R}^{m \times n}$ Distribuição símbolos observáveis



Parâmetros do Modelo

- $n \in \mathbb{N}$ Número de estados ocultos
- $m \in \mathbb{N}$ Número de símbolos observáveis
- $\Pi_0 \in \mathbb{R}^n$ Distribuição do estado inicial
- $O \in \mathbb{R}^{m \times n}$ Distribuição símbolos observáveis
- $T \in \mathbb{R}^{n \times n}$ Matriz transferência



Aplicações do Modelo

- Reconhecimento de Fala
  - Símbolos observáveis são espectros da Transformada de Fourier com filtros

Aplicações do Modelo

- Reconhecimento de Fala
  - Símbolos observáveis são espectros da Transformada de Fourier com filtros

- Modelagem de Volatilidade do PIB
  - Experimento com dados trimestrais do Japão, Reino Unido e Estados Unidos
  - Resultado: modelo passou nos testes de Hipótese Nula e Auto-correlação

Aprendizagem Espectral

- Motivação
- Definições iniciais
- Representação

Motivação

Principais perguntas:

$$P((o_i)_{i=1}^{t}), P(o_t \mid (o_i)_{i=1}^{t-1})$$

Motivação

Principais perguntas:

$$P((o_i)_{i=1}^{t}), P(o_t \mid (o_i)_{i=1}^{t-1})$$

Motivação

Principais perguntas:

$$P((o_i)_{i=1}^t), P(o_t \mid (o_i)_{i=1}^{t-1})$$

Ideia para responder a primeira pergunta

$$A_x = T\, diag((O_{x,i})_{i=1}^n)$$

$$P((o_i)_{i=1}^t) = \mathbf{1}^{\mathrm{T}} \prod_{i=1}^{t} A_{o_i} \Pi_0$$

Definições Iniciais

- Probabilidade do Símbolo i ser visto no tempo 1

$$P_1 \in \mathbb{R}^m$$

Definições Iniciais

- Probabilidade do Símbolo i ser visto no tempo 1

$$P_1 \in \mathbb{R}^m$$

- Probabilidade da dupla i, j ser vista nos tempos 1,2

$$P_{2,1} \in \mathbb{R}^{m \times m}$$

Definições Iniciais

- Probabilidade do Símbolo i ser visto no tempo 1

$$P_1 \in \mathbb{R}^m$$

- Probabilidade da dupla i, j ser vista nos tempos 1,2

$$P_{2,1} \in \mathbb{R}^{m \times m}$$

- Probabilidade da dupla i, j ser vista nos tempos 1,3 e x é vista na 2

$$P_{3,x,1} \in \mathbb{R}^{m \times m} \forall x \in [m]$$

Definições Iniciais

- Probabilidade do Símbolo i ser visto no tempo 1

$$P_1 \in \mathbb{R}^m$$

- Probabilidade da dupla i, j ser vista nos tempos 1,2

$$P_{2,1} \in \mathbb{R}^{m \times m}$$

- Probabilidade da dupla i, j ser vista nos tempos 1,3 e x é vista na 2

$$P_{3,x,1} \in \mathbb{R}^{m \times m} \forall x \in [m]$$

- Matriz esquerda da SVD reduzida de $P_{2,1}$

$$U \in \mathbb{R}^{m \times n}$$

Representação

$$b_1 = U^{\mathrm{T}} P_1 = (U^{\mathrm{T}} O)\Pi_0$$

$$b_\infty = (P_{2,1}^{\mathrm{T}} U)^+ P_1 = (U^{\mathrm{T}} O)^{-T} \mathbf{1}$$

$$B_x = (U^{\mathrm{T}} P_{3,x,1})(U^{\mathrm{T}} P_{2,1})^+ = (U^{\mathrm{T}} O) A_x (U^{\mathrm{T}} O)^{-1} \ \forall x \in [m]$$

Representação

$$b_1 = U^{\mathrm{T}} P_1 = (U^{\mathrm{T}} O)\Pi_0$$

$$b_\infty = (P_{2,1}^{\mathrm{T}} U)^{+} P_1 = (U^{\mathrm{T}} O)^{-T} \mathbf{1}$$

$$B_x = (U^{\mathrm{T}} P_{3,x,1})(U^{\mathrm{T}} P_{2,1})^{+} = (U^{\mathrm{T}} O) A_x (U^{\mathrm{T}} O)^{-1} \ \forall x \in [m]$$

Retornando à distribuição de probabilidade:

$$P((o_i)_{i=1}^{t}) = b_\infty^{\mathrm{T}} \prod_{i=1}^{t} B_{x_t} b_1 = \mathbf{1}^{\mathrm{T}} (U^{\mathrm{T}} O)^{-1} \left(\prod_{i=1}^{t} (U^{\mathrm{T}} O) A_{x_t} (U^{\mathrm{T}} O)^{-1} \right) (U^{\mathrm{T}} O)\Pi_0$$

Experimentos

- Hipótese
- Metodologia
- Resultados

Hipóteses

Da representação espectral era esperado:

- Tempo de treinamento menor
- Erro na probabilidade estimada maior

Metodologia

Métricas

- Média dos erros

$$\frac{\sum |\hat{p} - p|}{N}$$

Metodologia

Métricas

- Média dos erros

$$\frac{\sum |\hat{p} - p|}{N}$$

- Média dos erros normalizados

$$\frac{\sum |\hat{p} - p|^{1/l}}{N}$$

Metodologia

Métricas

- Média dos erros

$$\frac{\sum |\hat{p} - p|}{N}$$

- Média dos erros normalizados

$$\frac{\sum |\hat{p} - p|^{1/l}}{N}$$

- Tempo médio de treinamento

$$\frac{\sum t}{N}$$

Metodologia

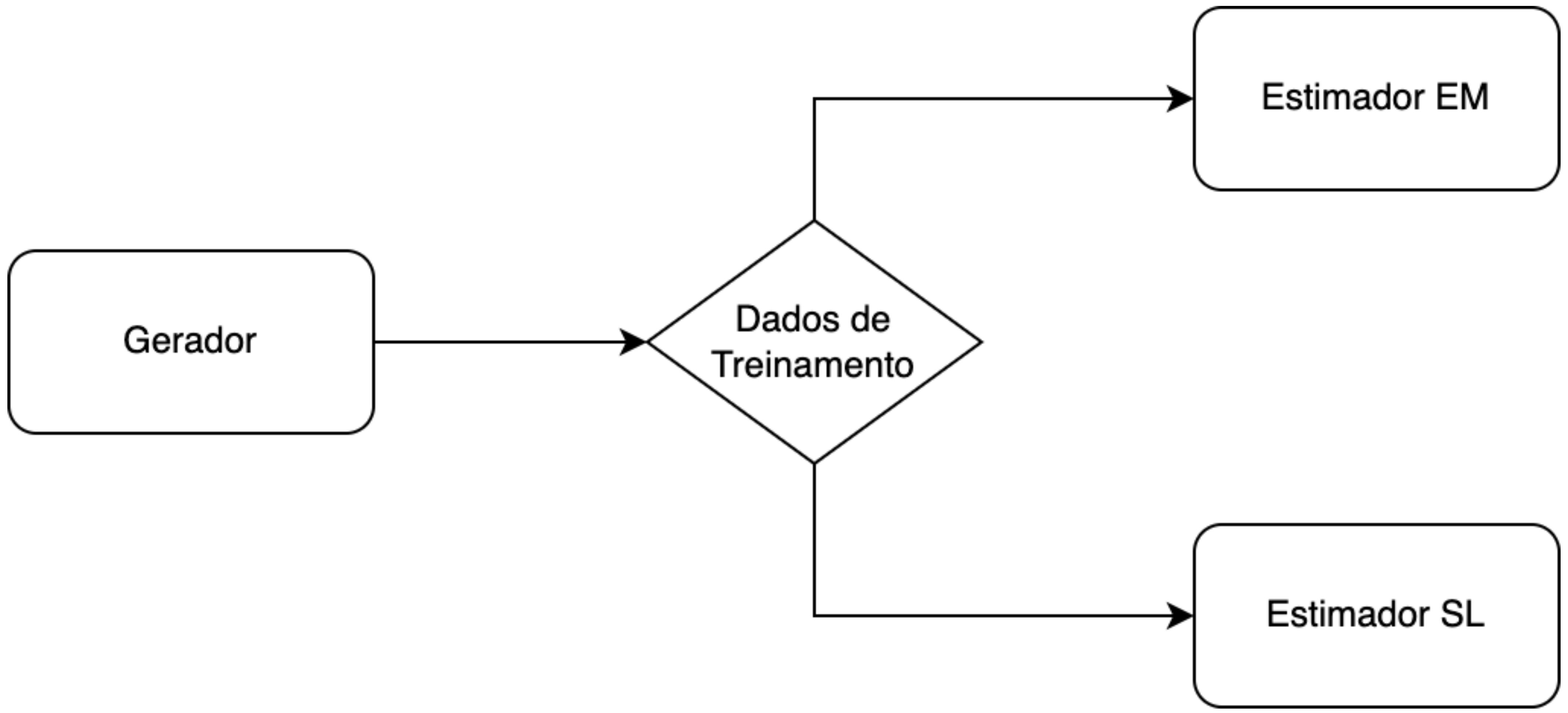
Configuração

- Modelos
  - Pequeno: 4 estados; 8 símbolos
  - Médio: 20 estados; 40 símbolos
  - Grande: 50 estados; 100 símbolos

Metodologia

Configuração

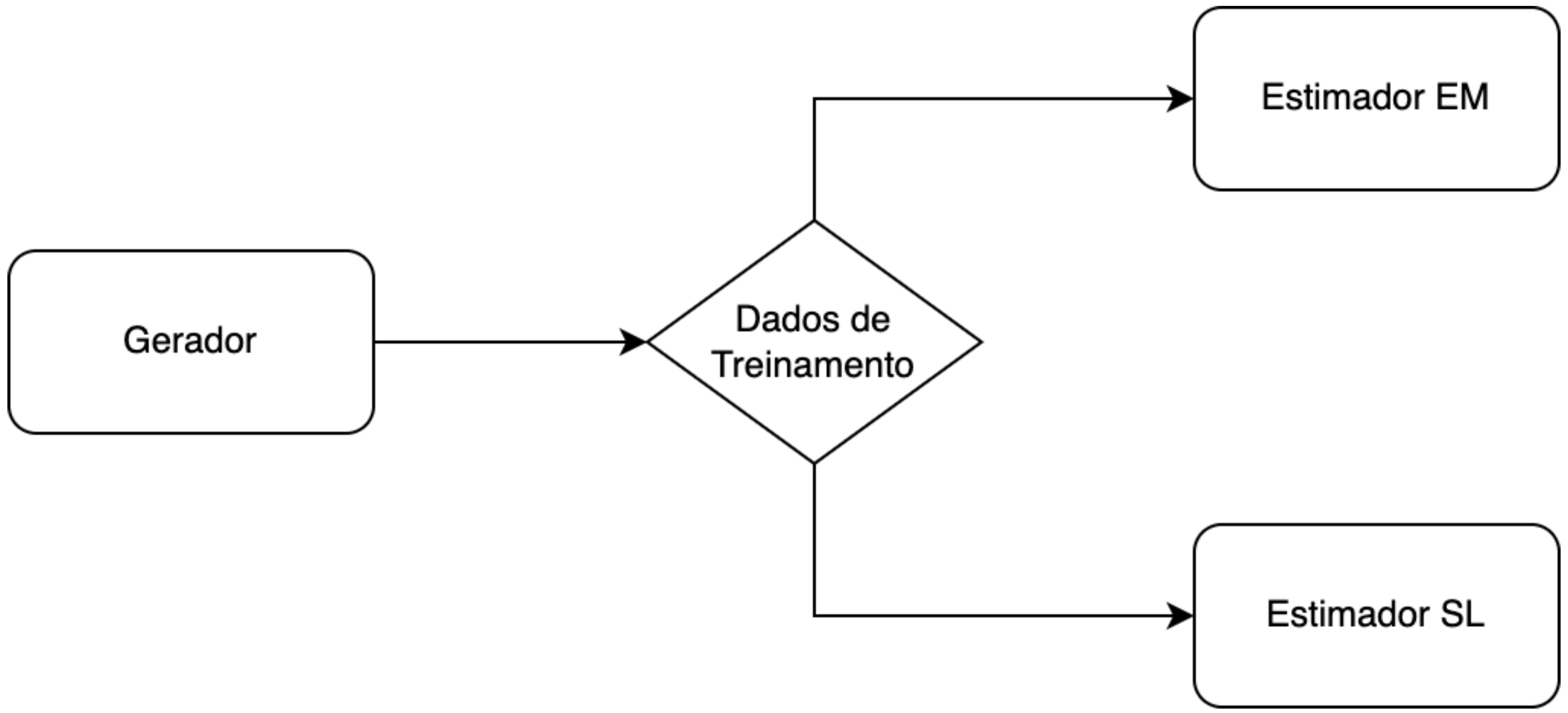
- Modelos
  - Pequeno: 4 estados; 8 símbolos
  - Médio: 20 estados; 40 símbolos
  - Grande: 50 estados; 100 símbolos

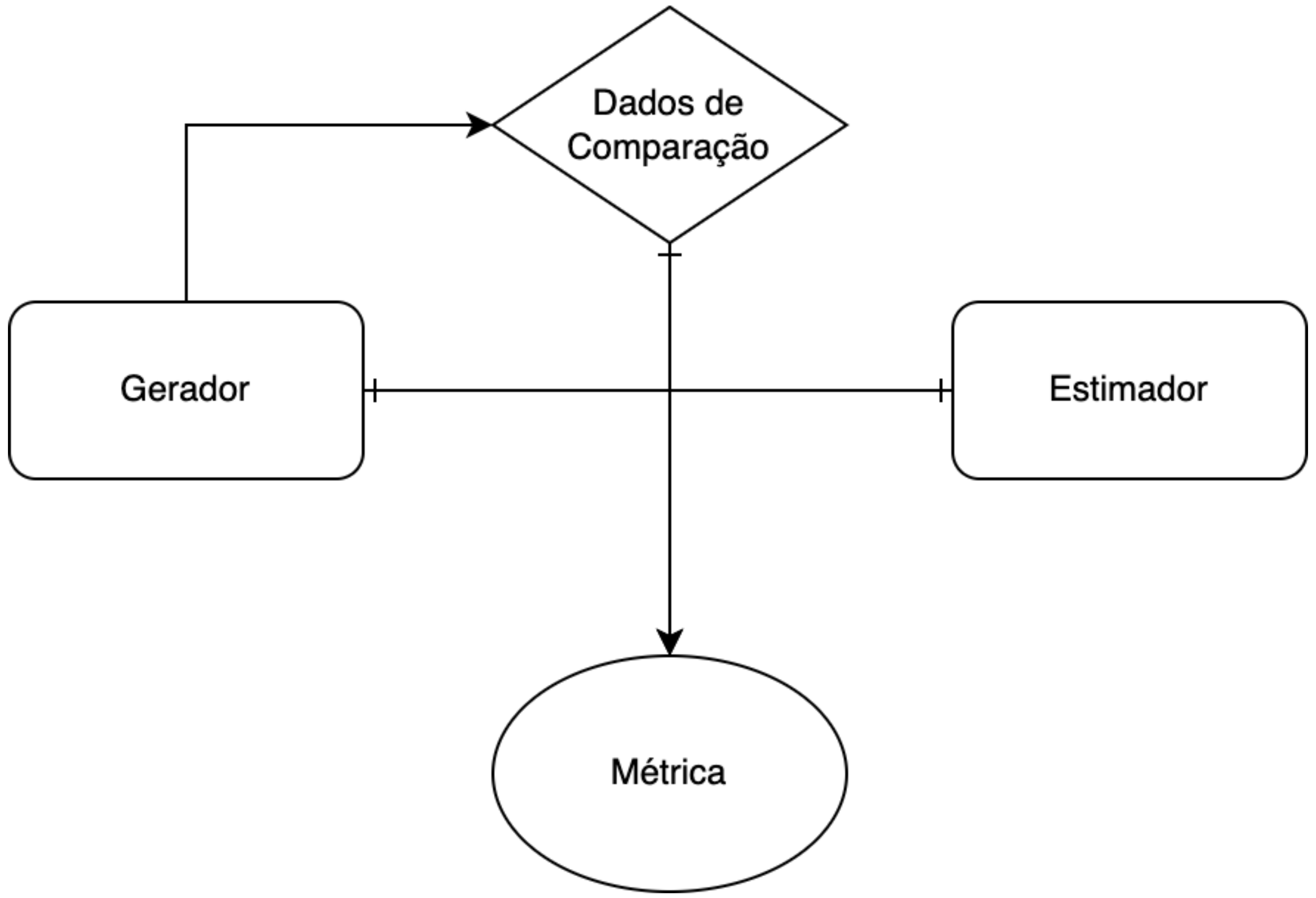


Metodologia

Configuração

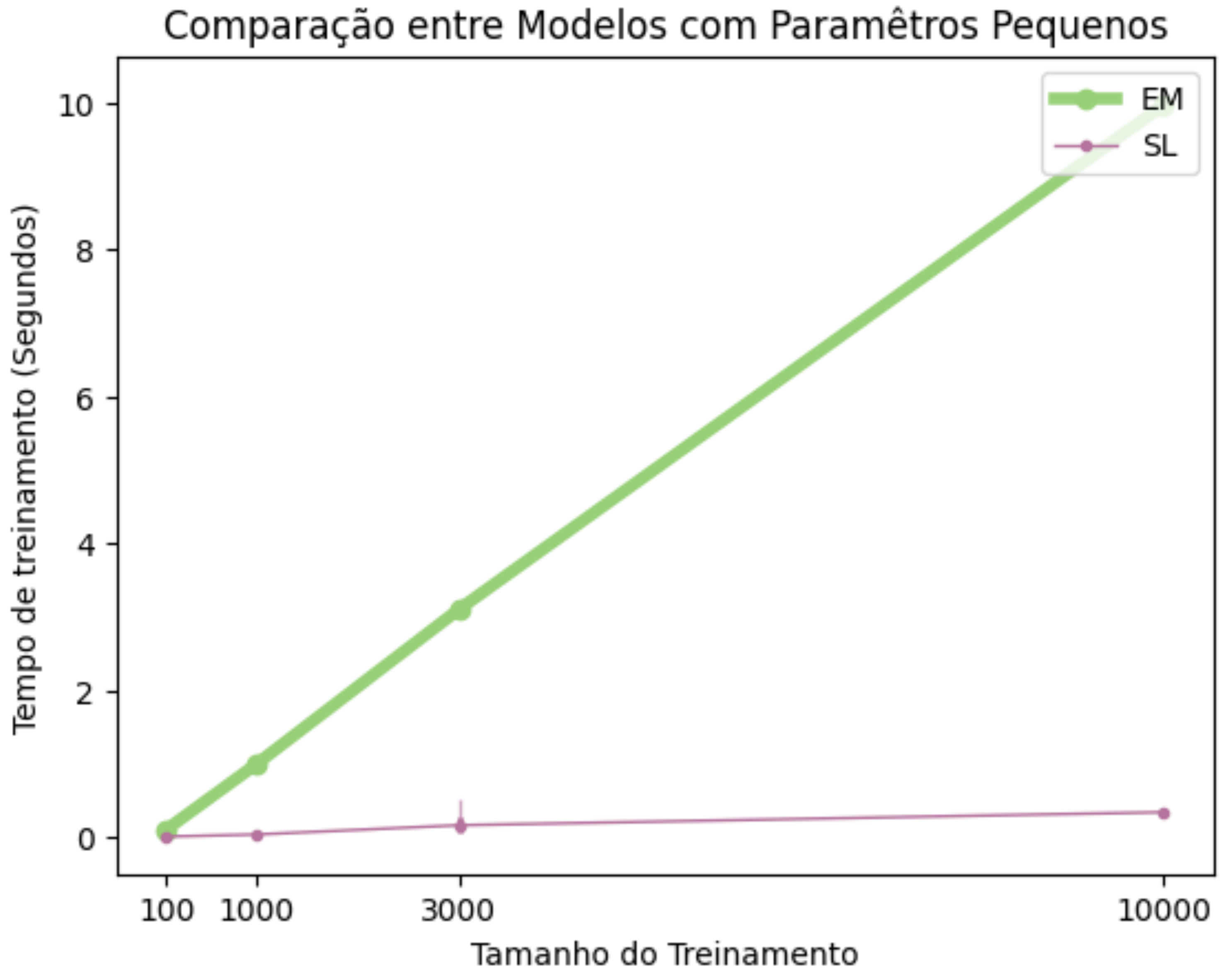
- Modelos
  - Pequeno: 4 estados; 8 símbolos
  - Médio: 20 estados; 40 símbolos
  - Grande: 50 estados; 100 símbolos

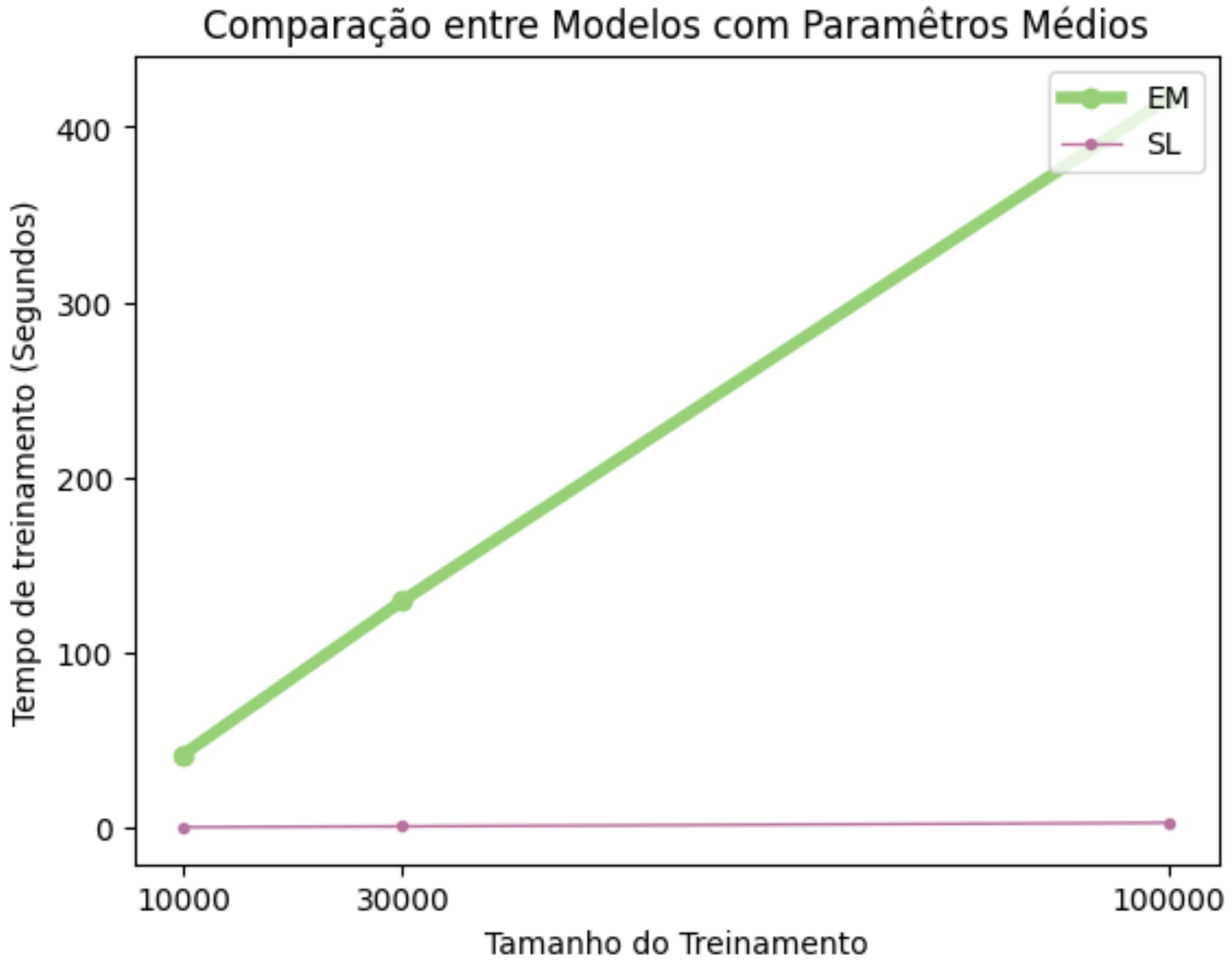


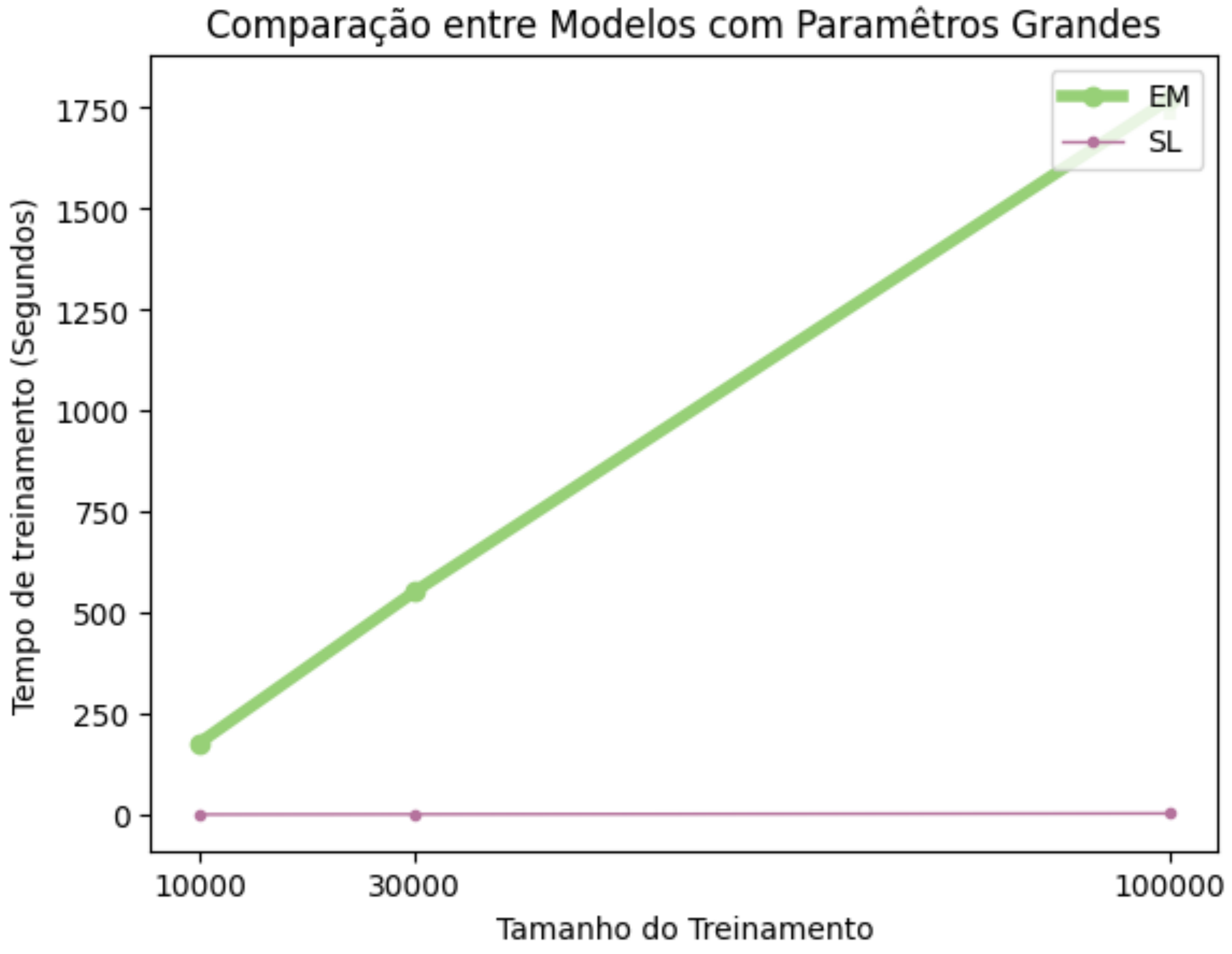


Resultados

Tempo

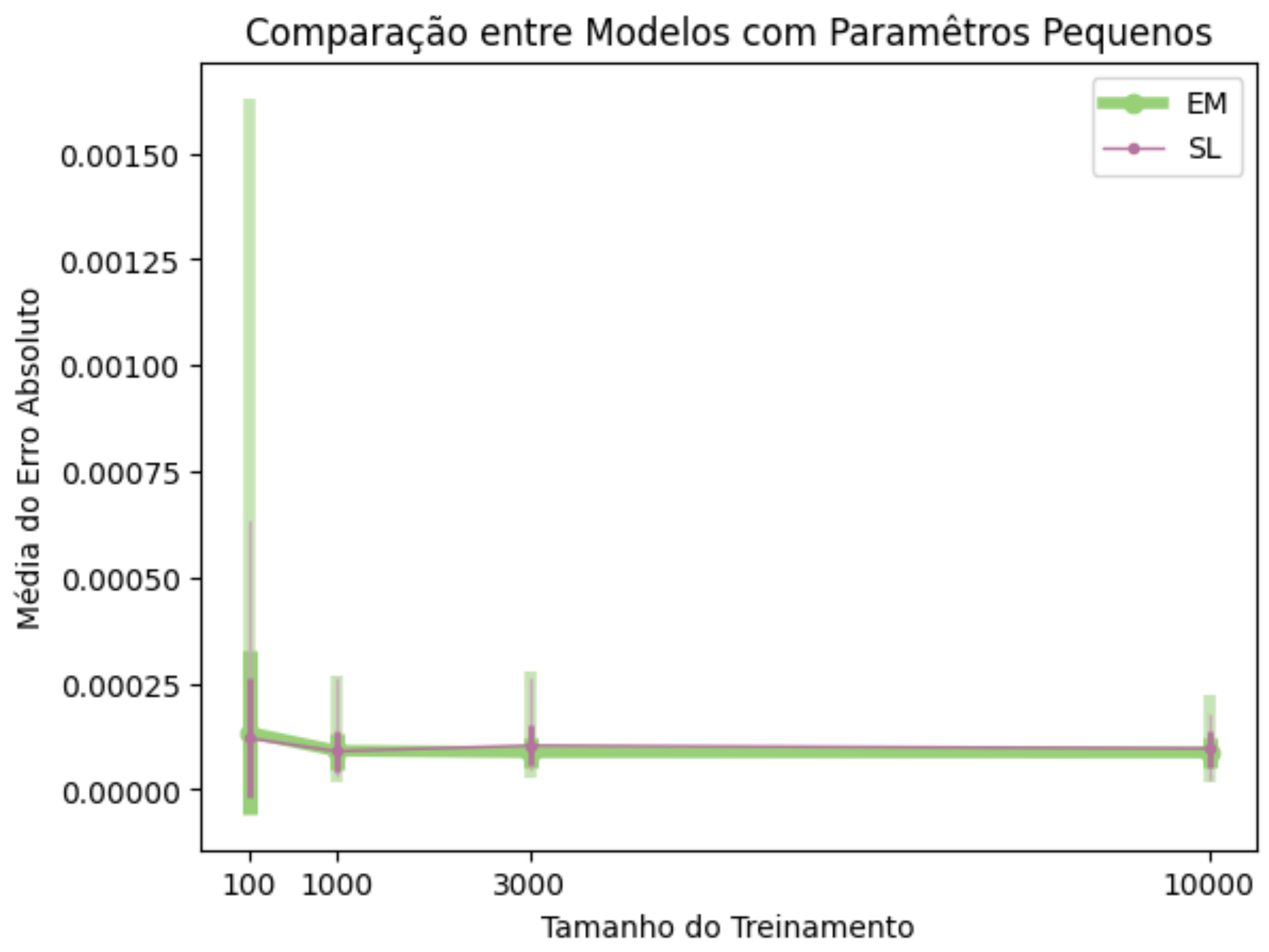


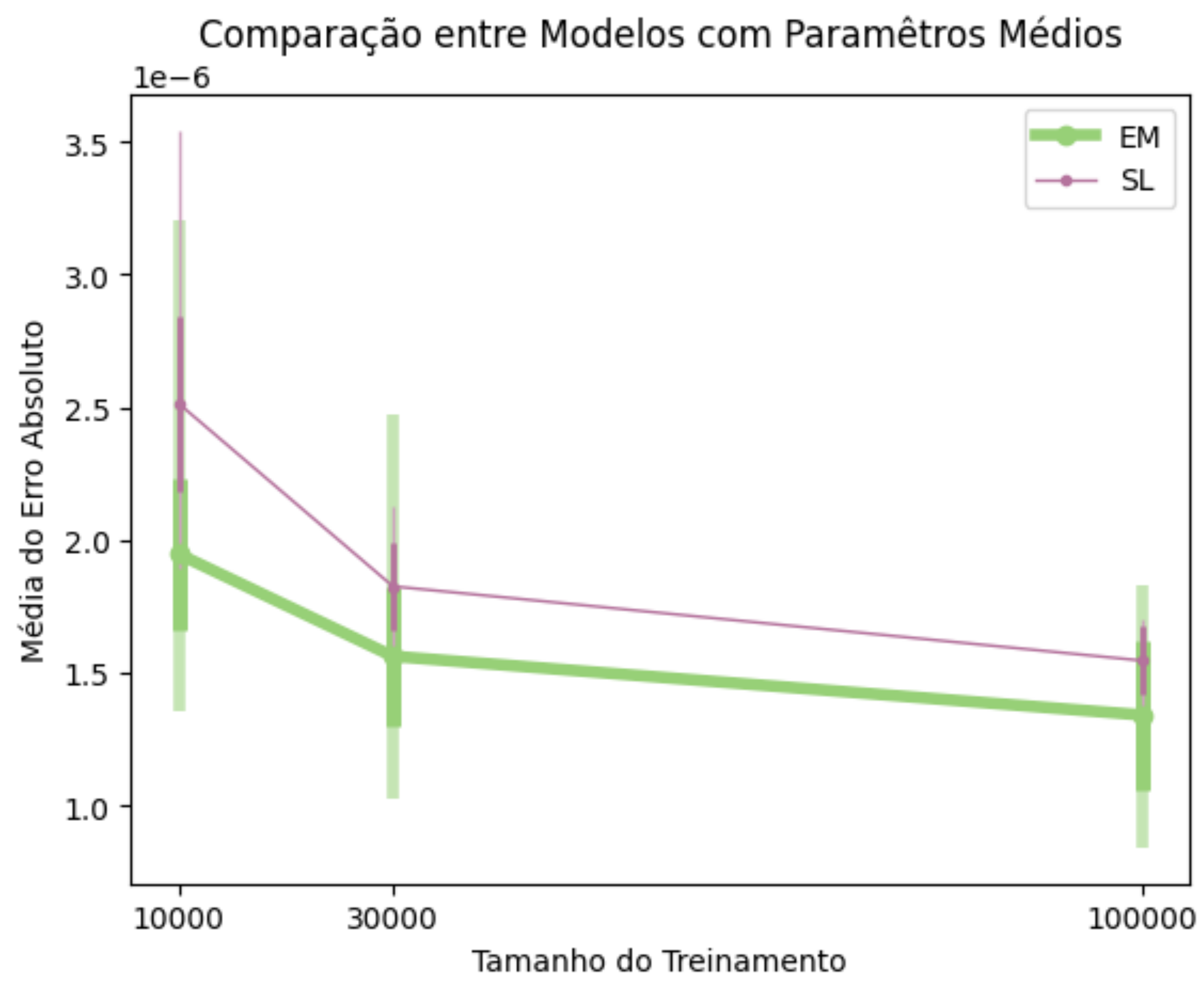


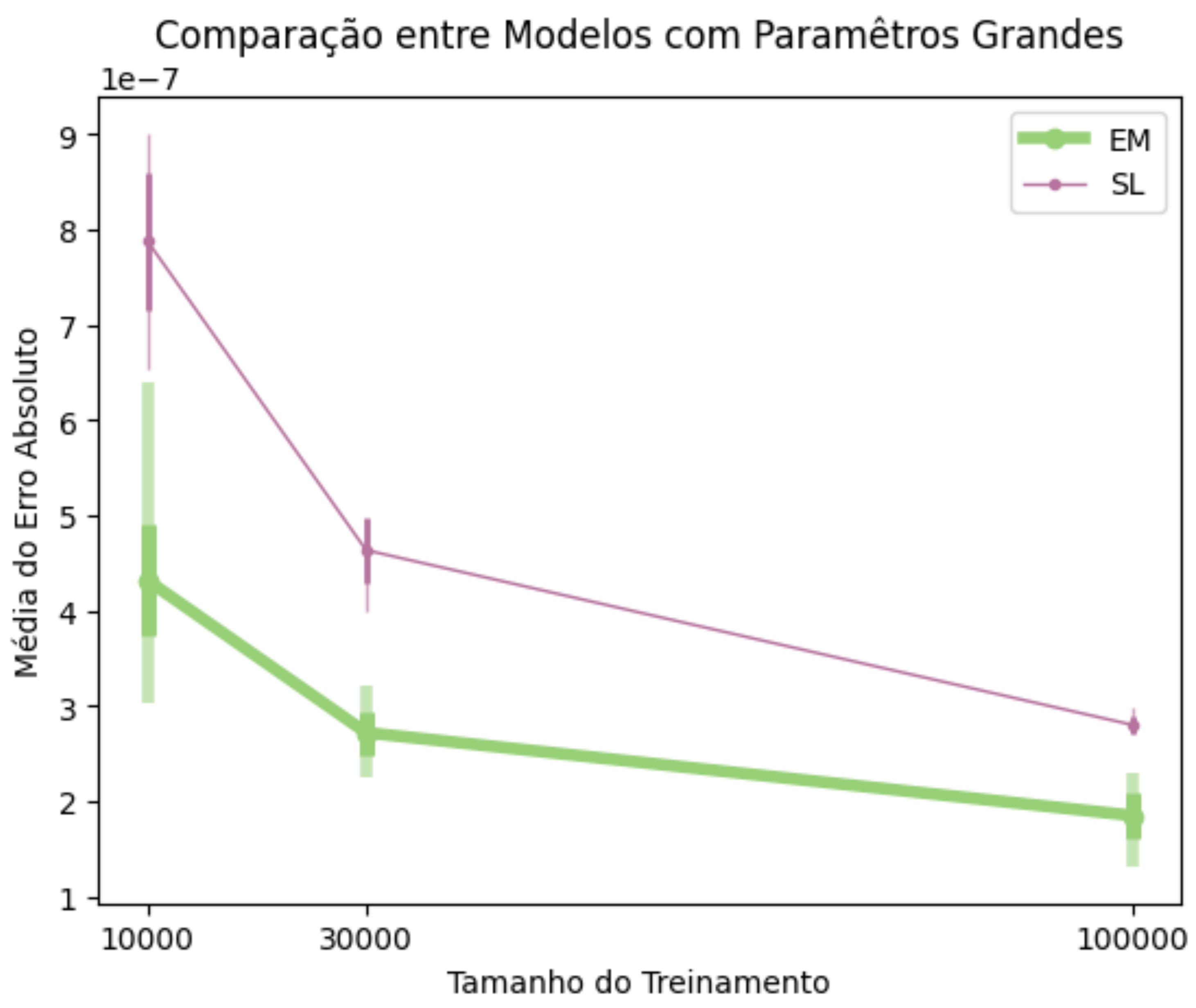


Resultados

Erro Médio

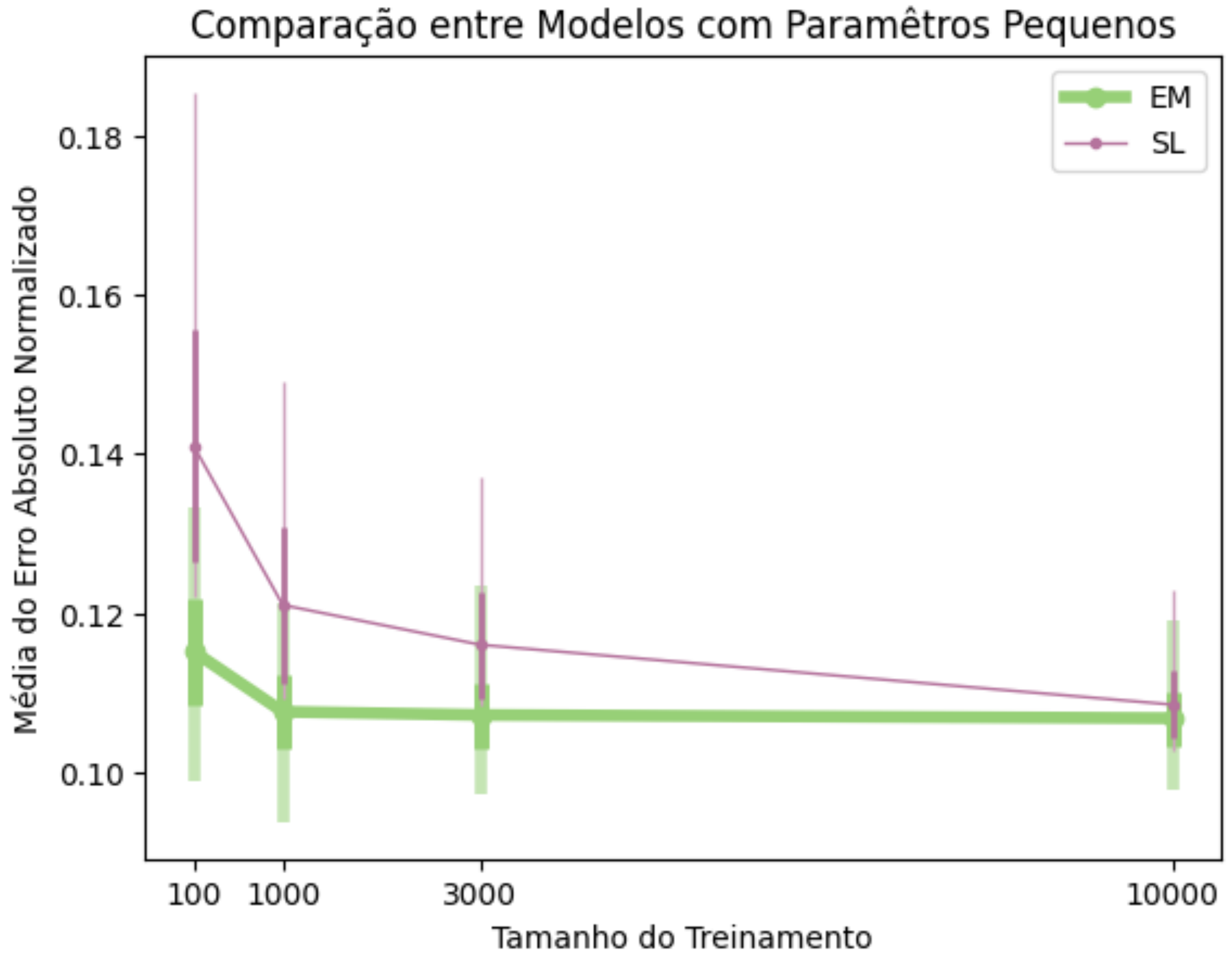


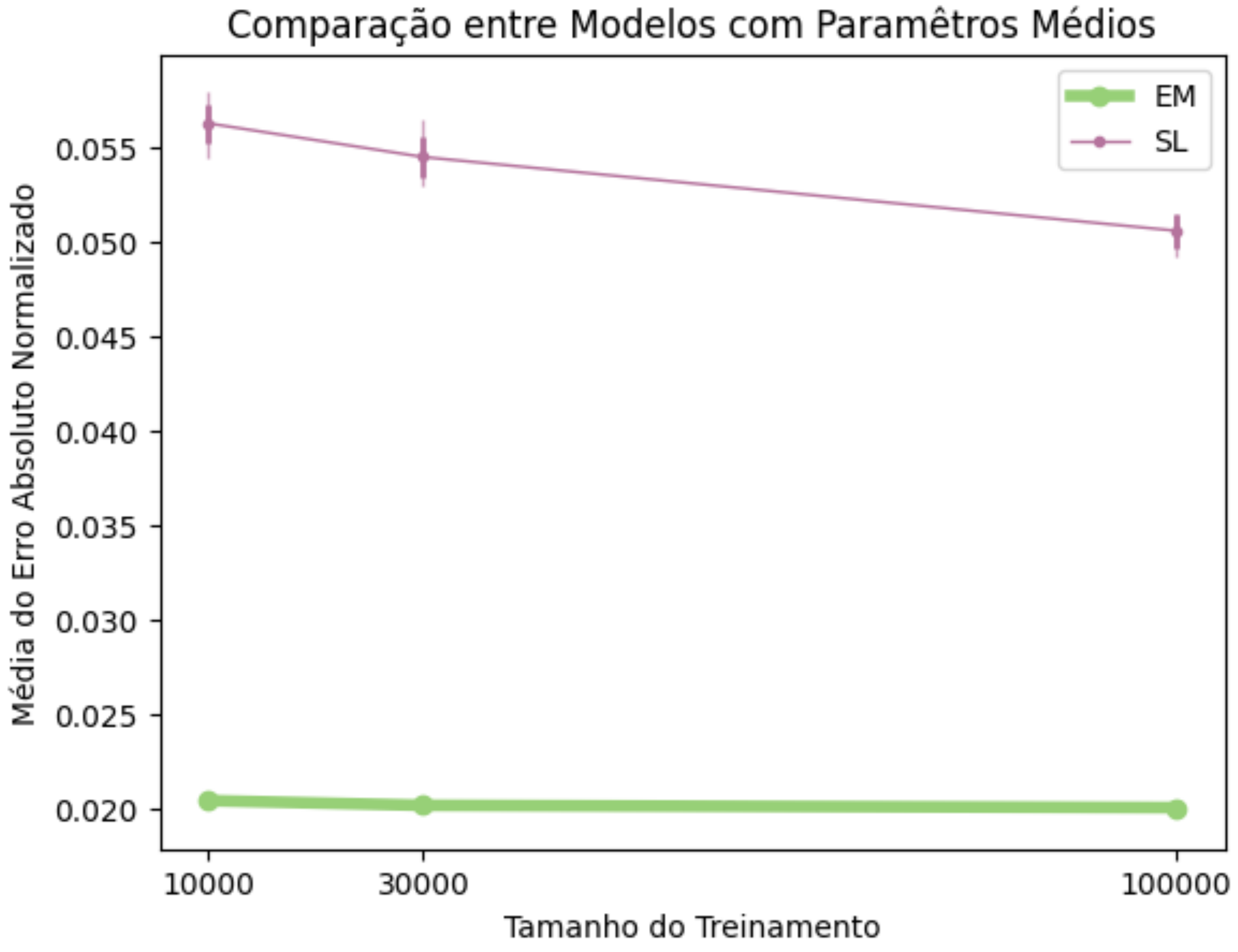


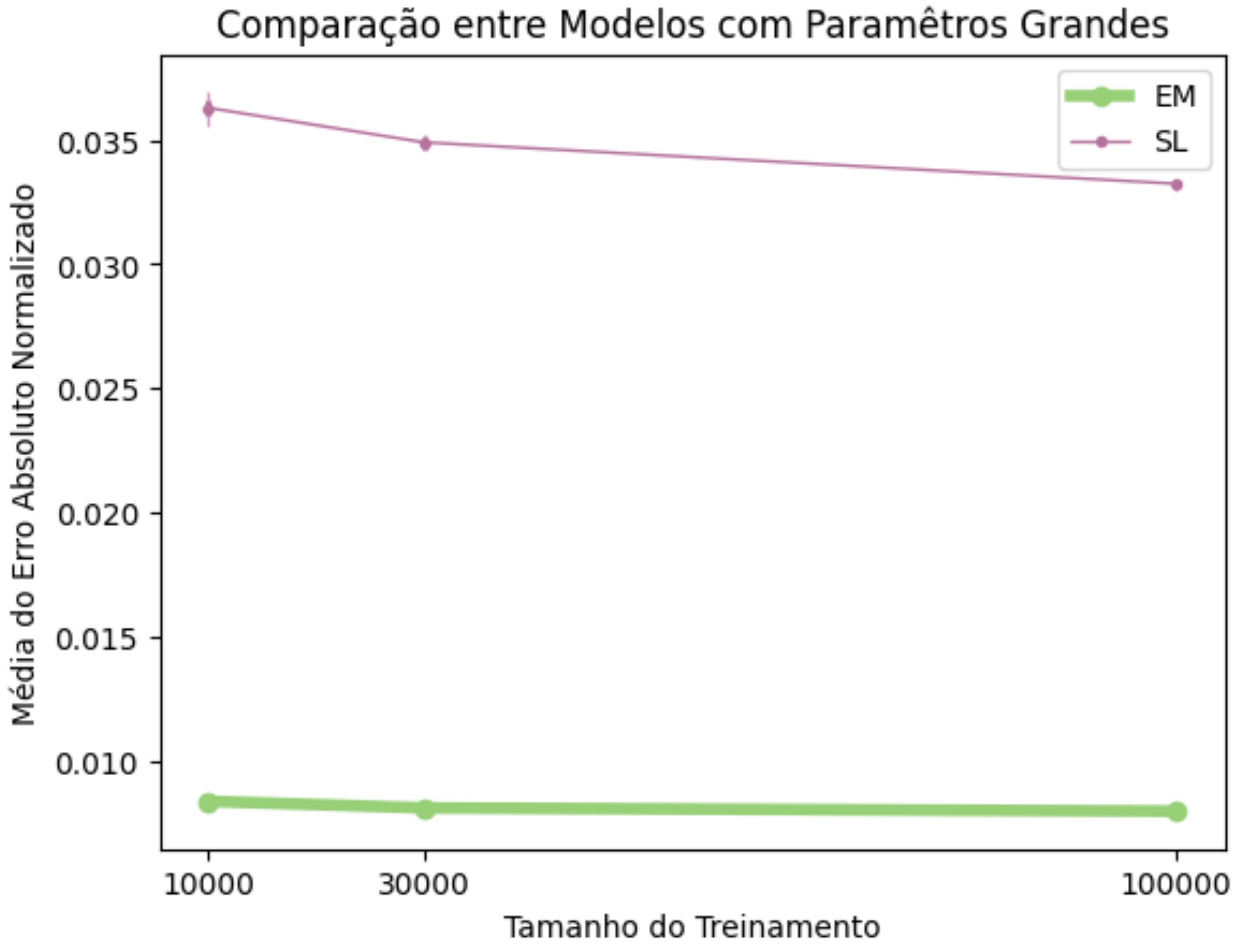


Resultados

Erro Médio Normalizado







Considerações Finais

- Probabilidades negativas

Considerações Finais

- Probabilidades negativas
- Comparação exclusivamente com triplas

Considerações Finais

- Probabilidades negativas
- Comparação exclusivamente com triplas
- Avaliar comportamento com processos que não atendem o critério $\Pi_0 > 0$

Considerações Finais

- Probabilidades negativas
- Comparação exclusivamente com triplas
- Avaliar comportamento com processos que não atendem o critério $\Pi_0 > 0$
- Analisar outros critérios de erro

Referências

- L.R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". Em: Proceedings of the IEEE 77.2 (1989), pp. 257–286. doi: 10.1109/5.18626.
- R. Bhar e S. Hamori. "Hidden Markov Models: Applications to Financial Economics". Springer New York, NY (2004), pp. 29-40. doi: 10.1007/b109046
- Daniel Hsu, Sham M. Kakade e Tong Zhang. "A spectral algorithm for learning Hidden Markov Models". Em: Journal of Computer and System Sciences 78.5 (2012). JCSS Special Issue: Cloud Computing 2011, pp. 1460–1480. issn: 0022-0000. doi: https://doi.org/10.1016/j.jcss.2011.12.025. url: https://www. sciencedirect.com/science/article/pii/S0022000012000244.