

PREVISÕES BAYESIANAS E MODELOS GRÁFICOS

Ralph dos Santos Silva

<http://www.ralphsilva.org/previsoesbayesianasemodelosgraficos.html>

Instituto de Matemática
Centro de Ciências Matemáticas e da Natureza
Universidade Federal do Rio de Janeiro

Ementa

- Modelos Gráficos Direcionados;
- Modelos Gráficos Não Direcionados; e
- Modelos em Espaço de Estados.

Referências

- *All of Statistics - A Concise Course in Statistical Inference*. Wasserman, L. (2010).
- *Bayesian Artificial Intelligence*. Korb, K. B. and Nicholson, A. E. (2011), 2nd Ed.
- *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Kjærulff, U. D. and Madsen, A. L. (2013), 2nd Ed.
- *Dynamic Linear Models with R*. Petris, G., Petrone, S. and Campagnoli, P. (2009).
- *Machine Learning - A Probabilistic Perspective*. Murphy, K. P. (2012).
- *Pattern Recognition and Machine Learning*. Bishop, C. M. (2006).

Modelo gráfico probabilístico

Fornece uma maneira simples de visualizar a estrutura de um modelo probabilístico e pode ser usado para se propor novos modelos.

Conhecimentos sobre a propriedade do modelo, incluindo a propriedade de independência condicional, podem ser obtidos pela inspeção do grafo.

Um **grafo** é uma estrutura abstrata que modela objetos e a relação (conexão) entre eles.

Cálculos complexos, necessários para realizar inferência e aprendizado em modelos sofisticados, podem ser expressos em termos de manipulações gráficas, na qual expressões matemáticas subjacentes são carregadas implicitamente.

Exemplo

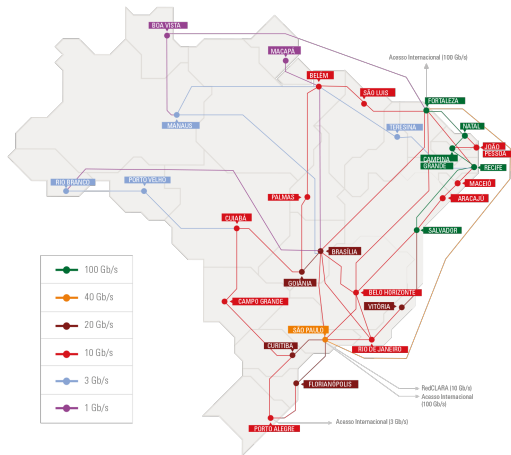


Figura 1: Mapa 2020 da Rede Ipê - Rede Nacional de Ensino e Pesquisa.

Fonte: <http://www.pop-go.rnp.br>

Exemplo

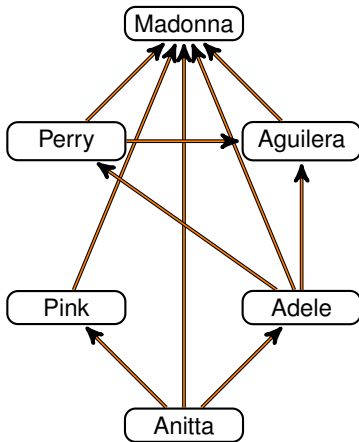


Figura 2: Uma pequena representação de uma rede social direcionada.

Grafo

Um **grafo** é uma estrutura abstrata que modela objetos e a relação (conexão) entre eles.



Figura 3: Um grafo direcionado (esquerda) e um grafo não direcionado (direita).

Um **grafo** é constituído de **vértices** (**nós**) conectados por **arestas** (**arcos**).

Neste curso, cada **vértice** representa uma **variável aleatória**.

Ademais, cada **aresta** expressa uma **relação probabilística** entre essas **variáveis aleatórias**.

Então, o **grafo** descreve como a distribuição conjunta das **variáveis aleatórias** pode ser decomposta.

Começaremos com os **modelos gráficos direcionados**¹.

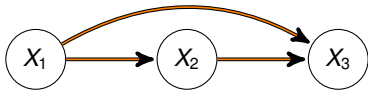


Figura 4: Um exemplo de **grafo direcionado**.

Depois, veremos os **modelos gráficos não direcionados**².

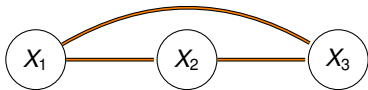


Figura 5: Um exemplo de **grafo não direcionado**.

¹Equivocadamente também chamados de **redes bayesianas**.

²Também conhecidos como **campos aleatórios de Markov**.

Grafo direcionado

Um grafo direcionado $\mathcal{G}$ consiste em um conjunto $\mathbf{X}$ de vértices e um conjunto $\mathbf{A}$ de arestas de pares ordenados de vértices: $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$.

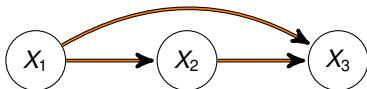


Figura 6: Grafo direcionado com $\mathbf{X} = (X_1, X_2, X_3)$ e $\mathbf{A} = [(X_1, X_2), (X_1, X_3), (X_2, X_3)]$.

Se $(X_1, X_2) \in \mathbf{A}$, então existe uma aresta (seta) saindo do vértice X_1 em direção ao vértice X_2 .

Se existir uma aresta saindo do vértice X_1 em direção ao vértice X_2 , então dizemos que

- X_1 e X_2 são adjacentes³; e
- X_1 é um pai de X_2 e X_2 é um filho de X_1 .

³Na verdade o sentido da aresta (seta) não importa e vale para grafos não direcionados.

Um **caminho direcionado** entre X_i e X_j é um conjunto de arestas (setas) ligando estes vértices.



Figura 7: **Caminho direcionado** entre X_1 e X_n .

Um **caminho não direcionado** entre X_i e X_j é um conjunto qualquer de arestas ligando vértices adjacentes entre eles.

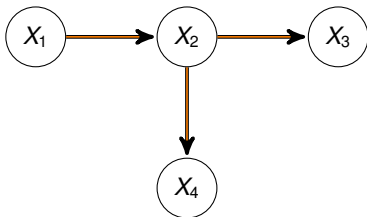


Figura 8: **Caminho não direcionado** entre X_3 e X_4 .

Um caminho direcionado que começa em X_i e termina na mesma variável X_i é chamado de **ciclo**.

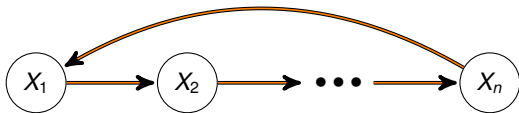


Figura 9: Grafo descrevendo um **ciclo** em X_1 .

Neste curso, os grafos direcionados estão sujeitos à seguinte importante restrição: **não deve existir ciclos diretos**.

Um grafo direcionado é chamado de **grafo acíclico direcionado** (DAG⁴, do inglês) se ele não possuir nenhum ciclo.

⁴ Directed Acyclic Graph

Independência condicional

Definição

Sejam X_1 , X_2 e X_3 variáveis aleatórias. X_1 e X_2 são condicionalmente independentes dado X_3 , denotado por $X_1 \perp X_2 | X_3$, se

$$f_{X_1, X_2 | X_3}(x_1, x_2 | x_3) = f_{X_1 | X_3}(x_1 | x_3) f_{X_2 | X_3}(x_2 | x_3).$$

Alternativamente, $X_1 \perp X_2 | X_3$ se

$$f_{X_1 | X_2, X_3}(x_1 | x_2, x_3) = f_{X_1 | X_3}(x_1 | x_3).$$

Daí, segue-se algumas propriedades:

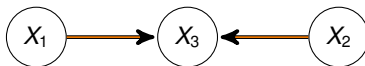
- Se $X_1 \perp X_2 | X_3$, então $X_2 \perp X_1 | X_3$;
- Se $X_1 \perp X_2 | X_3$ e $U = h(X_1)$, então $U \perp X_2 | X_3$;
- Se $X_1 \perp X_2 | X_3$ e $U = h(X_1)$, então $X_1 \perp X_2 | (X_3, U)$ ⁵;
- Se $X_1 \perp X_2 | X_4$ e $X_1 \perp X_3 | (X_2, X_4)$, então $X_1 \perp (X_2, X_3) | X_4$; e
- Se $X_1 \perp X_2 | X_3$ e $X_1 \perp X_3 | X_2$, então $X_1 \perp (X_2, X_3)$ ⁶.

⁵ $f_{X_2 | X_1, X_3, U}(x_2 | x_1, x_3, u) = f_{X_2 | X_3, U}(x_2 | x_3, u)$

⁶ Para eventos com probabilidade estritamente positivas.

Conexões

Convergente:

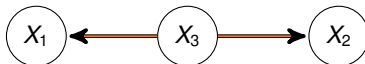


- X_1 e X_2 são marginalmente **independentes**:

$$X_1 \perp X_2 \Leftrightarrow f(x_1|x_2) = f(x_1) \text{ e } f(x_2|x_1) = f(x_2).$$

- X_1 e X_2 são condicionalmente (dado X_3) **dependentes**.

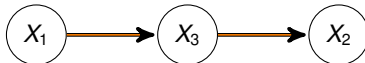
Divergente:



- X_1 e X_2 são marginalmente **dependentes**.
- X_1 e X_2 são condicionalmente (dado X_3) **independentes**:

$$X_1 \perp X_2|X_3 \Leftrightarrow f(x_1|x_2, x_3) = f(x_1|x_3) \text{ e } f(x_2|x_1, x_3) = f(x_2|x_3).$$

Série:



- X_1 e X_2 são marginalmente **dependentes**.
- X_1 e X_2 são condicionalmente (dado X_3) **independentes**:

$$X_1 \perp X_2|X_3 \Leftrightarrow f(x_1|x_2, x_3) = f(x_1|x_3) \text{ e } f(x_2|x_1, x_3) = f(x_2|x_3).$$

Modelos gráficos direcionados

Sejam X_1 , X_2 e X_3 variáveis aleatórias com uma distribuição conjunta dada por $f(x_1, x_2, x_3)$.

Podemos descrever a distribuição conjunta por

$$f(x_1, x_2, x_3) = f(x_3|x_1, x_2)f(x_1, x_2) = f(x_3|x_1, x_2)f(x_2|x_1)f(x_1).$$

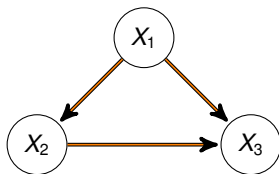


Figura 10: Um modelo gráfico direcionado que representa $f(x_3|x_1, x_2)f(x_2|x_1)f(x_1)$.

Agora, sejam X_1 , X_2 , X_3 e X_4 variáveis aleatórias tal que

$$f(x_1, x_2, x_3, x_4) = f(x_4|x_1, x_2, x_3)f(x_3|x_1, x_2)f(x_2|x_1)f(x_1).$$

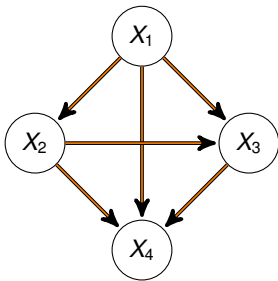


Figura 11: Um modelo gráfico direcionado de uma decomposição de $f(x_1, x_2, x_3, x_4)$.

Em geral, sejam $X_1, X_2, \dots, X_K$ variáveis aleatórias tal que

$$f(x_1, x_2, \dots, x_K) = f(x_K | x_1, x_2, \dots, x_{K-1}) \dots f(x_2 | x_1) f(x_1).$$

É possível representar esta decomposição através de um grafo com K vértices.

Cada vértice representa uma distribuição condicional da decomposição.

Note que o i -ésimo vértice X_i tem pais $\{X_1, X_2, \dots, X_{i-1}\}$, $i = 2, \dots, K$.

Dizemos que este grafo é **completamente conectado** porque existe uma aresta entre cada par de vértices.

Contudo, **é a ausência de algumas arestas no grafo que traz informações relevantes sobre a distribuição conjunta de $(X_1, X_2, \dots, X_K)$.**

Exemplo

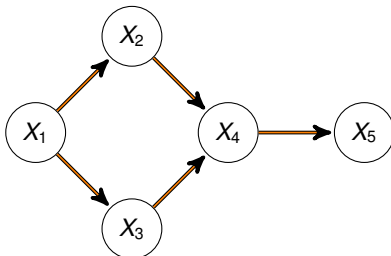


Figura 12: Um modelo gráfico direcionado acíclico.

Dado $\mathcal{G}$, a distribuição conjunta de $(X_1, X_2, \dots, X_5)$ é dada por

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_1)f(x_2|x_1)f(x_3|x_1)f(x_4|x_2, x_3)f(x_5|x_4).$$

Assim, podemos descrever algumas relações entre as variáveis aleatórias:

- $X_2 \perp X_3 | X_1$;
- $X_1 \perp X_4 | (X_2, X_3)$; e
- $(X_1, X_2, X_3) \perp X_5 | X_4$.

Exemplo

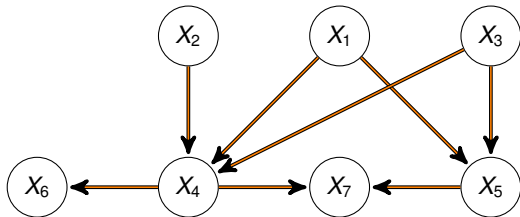


Figura 13: Um modelo gráfico direcionado acíclico.

Dado $\mathcal{G}$, a distribuição conjunta de $(X_1, X_2, \dots, X_7)$ é dada por

$$f(x_1, x_2, \dots, x_7) = f(x_1)f(x_2)f(x_3)f(x_4|x_1, x_2, x_3)f(x_5|x_1, x_3)f(x_6|x_4)f(x_7|x_4, x_5).$$

Assim, podemos descrever algumas relações entre as variáveis aleatórias:

- X_1, X_2 e X_3 são marginalmente independentes;
- $X_4 \perp X_5 | (X_1, X_2, X_3)$;
- X_4 e X_5 são condicionalmente (dado X_7) dependentes; e
- $X_6 \perp X_7 | (X_4, X_5)$.

Fatorização

Dado um grafo direcionado $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$ com $\mathbf{X} = (X_1, X_2, \dots, X_K)$, podemos descrever a distribuição conjunta das variáveis aleatórias por

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_K) = \prod_{k=1}^K f(x_k | \mathbf{p}_k),$$

em que $\mathbf{x} = (x_1, x_2, \dots, x_K)$ e $\mathbf{p}_k$ denota o conjunto formado pelos pais de x_k .

A seguir veremos um exemplo e introduziremos mais alguns elementos na modelagem gráfica.

Exemplo

Considere o modelo de **regressão linear** múltiplo dado por

$$Y_i = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_d z_{id} + \varepsilon_i, \quad \text{para } i = 1, 2, \dots, n,$$

com os ε_i independentes.

Além disso, por hipótese,

$$E(\varepsilon_i | \mathbf{z}_i) = 0 \quad \text{e} \quad \text{Var}(\varepsilon_i | \mathbf{z}_i) = \sigma^2.$$

Definimos os seguintes vetores:

$$\begin{aligned} \mathbf{z}_i &= (1, z_{i1}, z_{i2}, \dots, z_{id})^\top, \quad i = 1, 2, \dots, n, \\ \boldsymbol{\beta} &= (\beta_0, \beta_1, \dots, \beta_d)^\top. \end{aligned}$$

Em análise de regressão a estimação e demais procedimentos são **condicionais** a $\mathbf{z}_i$. Assim, temos $f(y_i | \boldsymbol{\beta}, \sigma^2, \mathbf{z}_i)$ para $i = 1, 2, \dots, n$.

Utilizaremos o **enfoque bayesiano**. Então, $\boldsymbol{\beta}$ e σ^2 são tratados como quantidades aleatórias. Por hipótese, **$\boldsymbol{\beta}$ e σ^2 são independentes a priori.**

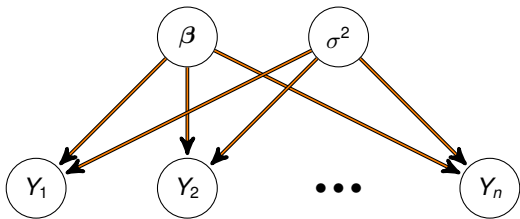


Figura 14: Construindo um modelo gráfico de regressão linear sob o enfoque bayesiano com β e σ^2 independentes a priori.

O grafo $\mathcal{R}$ acima tenta expressar a distribuição conjunta

$$f(\mathbf{y}, \beta, \sigma^2 | \mathbf{Z}) = f(\mathbf{y} | \beta, \sigma^2, \mathbf{Z}) f(\beta, \sigma^2) = \left[\prod_{i=1}^n f(y_i | \beta, \sigma^2, \mathbf{z}_i) \right] f(\beta) f(\sigma^2),$$

em que $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ e $\mathbf{Z} = (\mathbf{z}_1^\top, \mathbf{z}_2^\top, \dots, \mathbf{z}_n^\top)$ é matriz $n \times (d + 1)$.

$\mathbf{Z}$ e $\mathbf{z}_i$ indicam que não aparecem na modelagem gráfica (por enquanto).

Note que β e σ^2 compõem a distribuição de cada Y_i .

Simplificando a estrutura de um modelo gráfico

Em muitos modelos gráficos existem vértices que se conectam a diversos outros (veja a figura anterior).

A introdução de um retângulo (duplo e cantos arredondados) na representação gráfica simplifica a descrição e a visualização.

Neste elemento gráfico (retângulo) está descrito que os vértices β e σ^2 possuem arestas para cada vértice $Y_i, i = 1, 2, \dots, n$.

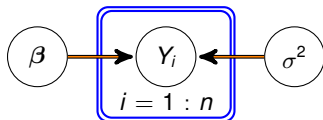


Figura 15: Introduzindo um retângulo (duplo e cantos arredondados) - representando diversas arestas - na construção de um modelo gráfico de regressão linear sob o enfoque bayesiano.

O grafo $\mathcal{R}$ acima expressa $f(\mathbf{y}, \beta, \sigma^2 | \mathbf{Z}) = \left[\prod_{i=1}^n f(y_i | \beta, \sigma^2, \mathbf{z}_i) \right] f(\beta) f(\sigma^2)$.

Quantidades determinísticas na estrutura de um modelo gráfico

Em modelos de regressão, a análise é condicional ao conhecimento de $\mathbf{z}_i$, $i = 1, 2, \dots, n$. Então, $\mathbf{z}_i$ **não é tratado como vetor aleatório (observado)**.

Às vezes é necessário representar quantidades determinísticas na modelagem gráfica. Isso é feito via um retângulo simples.

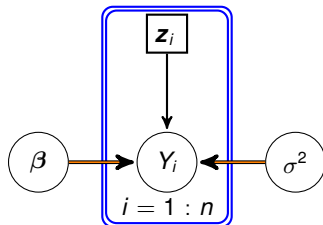


Figura 16: Introduzindo um retângulo simples (preto) - representando quantidades determinísticas - na construção de um modelo gráfico de regressão linear sob o enfoque bayesiano.

O grafo $\mathcal{R}$ acima expressa $f(\mathbf{y}, \beta, \sigma^2 | \mathbf{Z}) = \left[\prod_{i=1}^n f(y_i | \beta, \sigma^2, \mathbf{z}_i) \right] f(\beta) f(\sigma^2)$.

Distribuição a posteriori

Dada a amostra $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$, estamos interessados na distribuição a posteriori de β e σ^2 , isto é, $f(\beta, \sigma^2 | \mathbf{y}, \mathbf{Z})$.

Na modelagem gráfica preenchemos com alguma cor os círculos que representam os vértices observados.

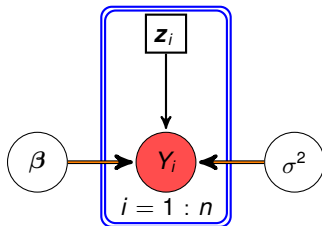


Figura 17: Introduzindo vértices preenchidos (vermelho) - representando as variáveis observadas - na construção de um modelo gráfico de regressão linear sob o enfoque bayesiano.

Do grafo $\mathcal{R}$ acima, temos $\left[\prod_{i=1}^n f(y_i | \beta, \sigma^2, \mathbf{z}_i) \right] f(\beta) f(\sigma^2) \propto f(\beta, \sigma^2 | \mathbf{y}, \mathbf{Z})$.

Elementos da modelagem

- Distribuição a priori:

$$f(\beta, \sigma^2) = f(\beta)f(\sigma^2) \text{ com } \beta \sim \mathcal{N}_d(\mathbf{0}; 10^6 \mathbf{I}) \text{ e } \sigma^2 \sim \mathcal{IG}(10^{-3}; 10^{-3}).$$

- Função de verossimilhança:

$$L(\beta, \sigma^2) = f(\mathbf{y}|\beta, \sigma^2, \mathbf{Z}) = \prod_{i=1}^n f(y_i|\beta, \sigma^2, \mathbf{z}_i).$$

- Distribuição a posteriori:

$$\begin{aligned} f(\beta, \sigma^2|\mathbf{y}, \mathbf{Z}) &= \frac{L(\beta, \sigma^2)f(\beta, \sigma^2)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\beta, \sigma^2, \mathbf{Z})f(\beta, \sigma^2)}{f(\mathbf{y})} \\ &\propto \left[\prod_{i=1}^n f(y_i|\beta, \sigma^2, \mathbf{z}_i) \right] f(\beta, \sigma^2). \end{aligned}$$

- Distribuição preditiva (verossimilhança marginal):

$$f(\mathbf{y}) = \int \left[\prod_{i=1}^n f(y_i|\beta, \sigma^2, \mathbf{z}_i) \right] f(\beta, \sigma^2) d\beta d\sigma^2.$$

Previsão

Em termos de modelagem bayesiana, a previsão para um determinado Y^* dado $\mathbf{z}^*$ é baseada na seguinte distribuição:

$$f(y^*|\mathbf{z}^*, \mathbf{y}, \mathbf{Z}) = \int f(y^*|\beta, \sigma^2, \mathbf{z}^*)f(\beta, \sigma^2|\mathbf{y}, \mathbf{Z})d\beta d\sigma^2.$$

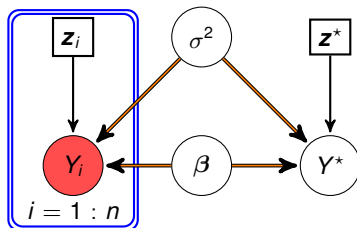


Figura 18: Introduzindo o vértice Y^* representando um valor desconhecido que desejamos prever dado $\mathbf{z}^*$ no modelo gráfico de regressão linear sob o enfoque bayesiano.

A previsão para a coleção Y_k^* dado $\mathbf{z}_k^*$, $k = 1, 2, \dots, m$, é dada pela seguinte distribuição:

$$f(\mathbf{y}^* | \mathbf{Z}^*, \mathbf{y}, \mathbf{Z}) = \int \left[\prod_{k=1}^m f(y_k^* | \beta, \sigma^2, \mathbf{z}_k^*) \right] f(\beta, \sigma^2 | \mathbf{y}, \mathbf{Z}) d\beta d\sigma^2,$$

com $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_m^*)^\top$ e $\mathbf{Z}^* = (\mathbf{z}_1^*, \mathbf{z}_2^*, \dots, \mathbf{z}_m^*)^\top$ é matriz $m \times (d + 1)$.

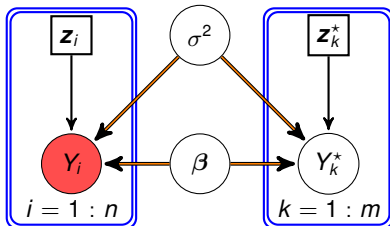


Figura 19: Introduzindo uma coleção de vértices Y_k^* , $k = 1, 2, \dots, m$, representando valores desconhecido que desejamos prever dado o respectivo $\mathbf{z}_k^*$ no modelo gráfico de regressão linear sob o enfoque bayesiano.