ACH2053 – Introdução à Estatística

Aula 09c: Otimização

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2025

Estimadores

O estimador:

$$\hat{\theta} = \arg\max_{\theta \in \Omega} f_n(\theta | x_1, \ldots, x_n)$$

é chamado de estimador bayesiano Maximum a Posteriori (MAP).

O estimador:

$$\hat{\theta} = \arg\max_{\theta \in \Omega} f_n(x_1, \ldots, x_n | \theta)$$

é chamado de máxima verossimilhança (M.L.E. - maximum likelihood estimator).

Estimador de Máxima Verossimilhança

Suponha que as n variáveis aleatórias $X_1, \ldots, X_n$ formam uma amostra aleatória de uma distribuição para qual a p.d.f. (p.m.f.) é $f(X|\theta)$. Então:

$$L(\theta; \mathbf{x}) = f(x_1|\theta) \cdots f(x_n|\theta).$$

Para encontrar o M.L.E. usualmente considera-se a transformação $\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x})$ e resolve-se a seguinte equação:

$$\nabla_\theta \log L(\theta) = 0.$$

Exemplo: Suponha que uma lâmpada regular, uma lâmpada de vida longa e uma lâmpada de vida extra-longa serão testadas. O tempo de vida de cada lâmpada é respectivamente X_1, X_2 e X_3 e possuem distribuição exponencial com média θ, 2θ e 3θ respectivamente.

Determine o M.L.E. de θ baseado nas observações X_1, X_2 e X_3.

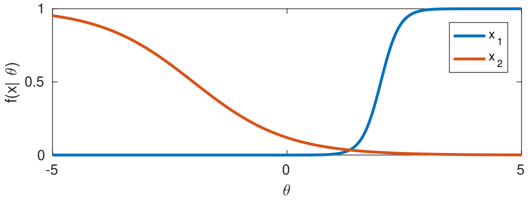
Considere o modelo ML2 para a Teoria de Resposta ao Item (TRI) para descrever a probabilidade de um aluno j acertar a questão i:

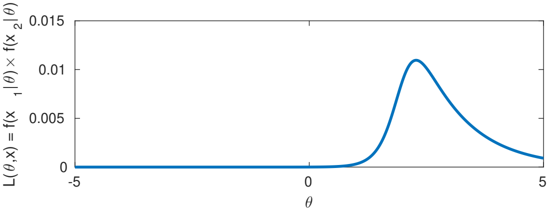
$$\Pr(A_{ij} = 1) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}},$$

onde $A_{ij} \in \{0,\ 1\}$ é a variável aleatória que indica se o aluno j acertou $(A_{ij} = 1)$ ou errou $(A_{ij} = 0)$ a questão i, θ_j representa a habilidade do aluno j, a_i representa o parâmetro de discriminação da questão i, e b_i representa o parâmetro de dificuldade da questão i.

Exemplo com duas questões

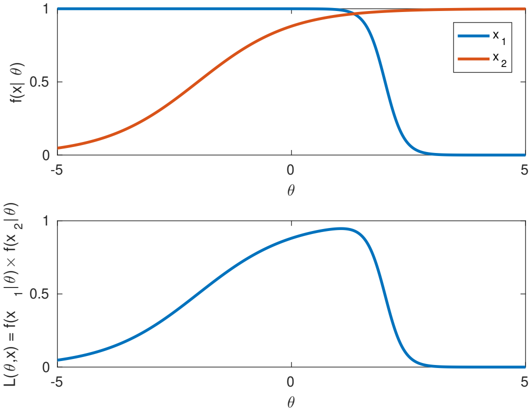
questão	a_1	b_1	x_1
Q1	5	2	1
Q2	1	-2	0





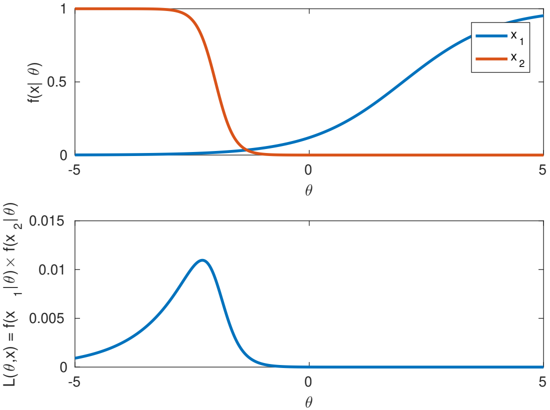
Exemplo com duas questões

questão	a_1	b_1	x_1
Q1	5	2	0
Q2	1	-2	1



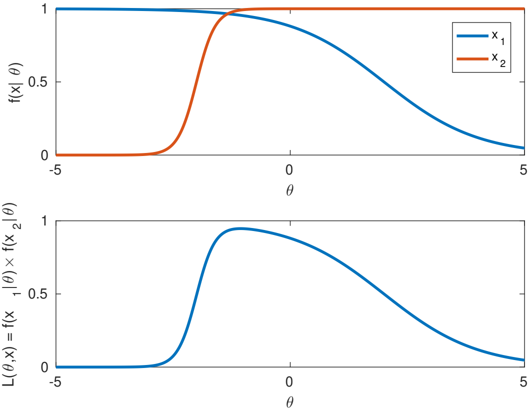
Exemplo com duas questões

questão	a_1	b_1	x_1
Q1	1	2	1
Q2	5	-2	0

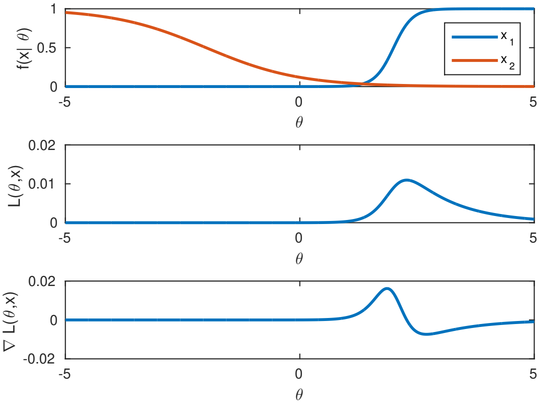


Exemplo com duas questões

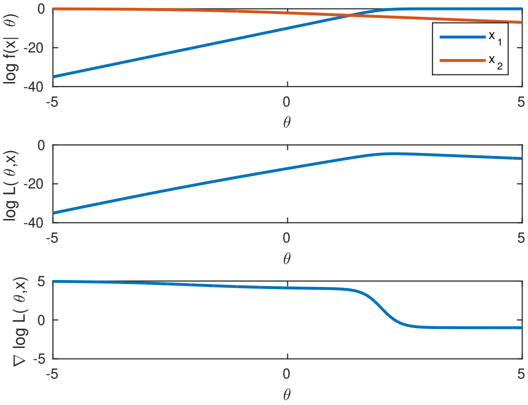
questão	a_1	b_1	x_1
Q1	1	2	0
Q2	5	-2	1



Maximizando Likelihood



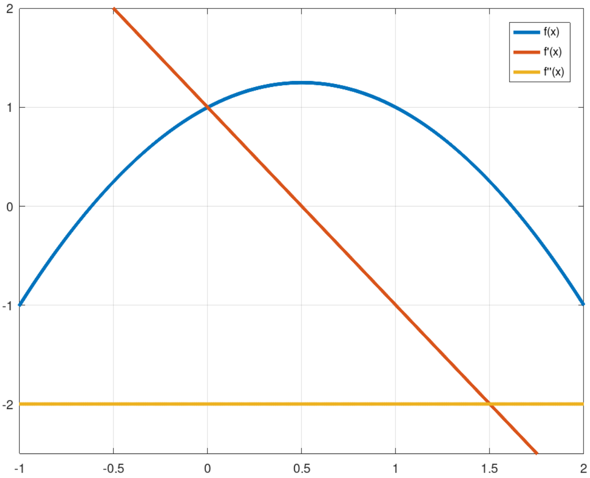
Maximizando log-Likelihood



Exemplos

Exercício 1: encontro o ponto máximo para a função:

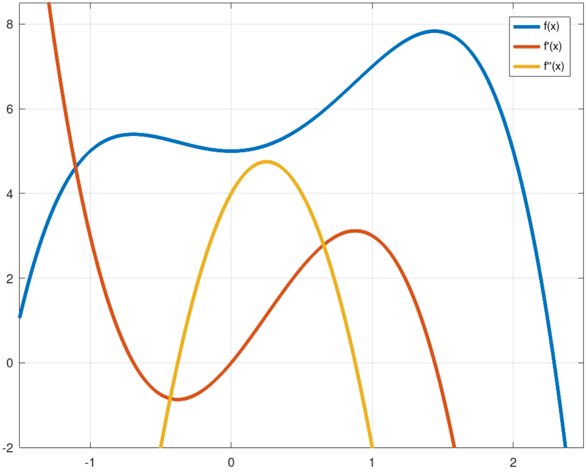
$$f(x) = -x^2 + x + 1$$



Exemplos

Exercício 2: encontro o ponto máximo para a função:

$$f(x) = -x^4 + x^3 + 2x^2 + 5$$



Problema de Otimização
Dada uma função $g : \mathbb{R}^d \to \mathbb{R}$, encontre $\mathbf{x}^* \in \mathbb{R}^d$ tal que $g(\mathbf{x}^*) \geq g(\mathbf{x})$ para todo $\mathbf{x} \in \mathbb{R}$.

Teorema

Se $g : \mathbb{R}^d \to \mathbb{R}$ é contínua e diferenciável, a solução $\mathbf{x}^*$ para o problema de otimização deve atender a seguinte equação:

$$\frac{\partial g(\mathbf{x}^*)}{\partial x_i} = g'(\mathbf{x}^*) = 0 \Leftrightarrow \nabla_{\mathbf{x}} g(\mathbf{x}^*) = \mathbf{0}.$$

onde x_i representa a i-ésima dimensão da entrada $\mathbf{x}$.

Método de Newton

Teorema (Série de Taylor)

Seja $f : \mathbb{R} \to \mathbb{R}$ uma função infinitamente diferenciável definida em um intervalo aberto $(a - r, a + r)$, então:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n,$$

onde $f^{(n)(a)}$ é a n-ésima derivada de f no ponto a.

Definition (Método de Newton)

Dada uma função $g : \mathbb{R} \to \mathbb{R}$ contínua e duas vezes diferenciáveis tal que $g''(x^*) \neq 0$, o método de Newton itera em valores $x^{(t)}$ seguindo:

$$x^{(t+1)} = x^{(t)} - \frac{g'(x^{(t)})}{g''(x^{(t)})}.$$

Método do Gradiente Ascendente

Definition (Escalar)

Dada uma função $g : \mathbb{R} \to \mathbb{R}$ contínua e diferenciável. O método do gradiente Ascendente itera nos valores $x^{(t)}$ seguindo:

$$x^{(t+1)} = x^{(t)} + \beta^{(t)} g'(x).$$

Definition (Vetorial)

Dada uma função $g : \mathbb{R}^d \to \mathbb{R}$ contínua e diferenciável. O método do gradiente Ascendente itera nos valores $\mathbf{x}^{(t)}$ seguindo:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \beta^{(t)} \nabla_{\mathbf{x}} g(\mathbf{x}^{(t)}).$$

Usualmente $\beta^{(t)} \to 0$.

Método do Gradiente Ascendente

1. Escolha $x^{(0)}$ arbitrário
2. Escolha $\beta^{(0)} > 0$ arbitrário
3. Enquanto não atende critério de parada
   3.1 Faça:
   $$\mathbf{x}^{(t+1)} \leftarrow \mathbf{x}^{(t)} + \beta^{(t)} \nabla_{\mathbf{x}} g\left(\mathbf{x}^{(t)}\right)$$
   3.2 Se: $g\left(\mathbf{x}^{(t+1)}\right) > g\left(\mathbf{x}^{(t)}\right)$
      (a) Então:
      $$\beta^{(t+1)} \leftarrow r\beta^{(t)}$$
      (b) Caso contrário:
      $$\beta^{(t+1)} \leftarrow \frac{1}{r}\beta^{(t)}$$
      $$\mathbf{x}^{(t+1)} \leftarrow \mathbf{x}^{(t)}$$

Critérios de Parada: gradiente mínimo, quantidade de iterações máximas
Busca de β: $r > 1$, mas existem vários outros métodos