

Análise Numérica

Raízes e otimização - 03

Renato Martins Assunção

DCC - UFMG

2012

Minimização de $g(x_1, \dots, x_m)$

- Suponha que $x = (x_1, x_2, x_3, \dots, x_m)$.
 - Queremos encontrar o ponto x que minimize $g(x)$.
 - Como fazer?
 - Ache os pontos críticos: gradiente $\nabla g = 0$
 - Como achar os pontos críticos?
 - $\nabla g = 0$ é um SISTEMA de equações não-lineares.
 - Use Newton multivariado : Chame $f = \nabla g$ e use nossa equação anterior.
- $$x_{n+1} = x_n - (\nabla^2 g)^{-1} * \nabla g$$
- Onde $\nabla^2 g$ é matriz Hessiana $m \times m$ com as derivadas parciais de segunda ordem.

Exemplo com dimensão 3

- Suponha que temos uma função de três variáveis: $g(b, a, c)$
- Começamos com um vetor inicial (b_0, a_0, c_0)
- E iteramos de acordo com a equação abaixo com as derivadas avaliadas no vetor corrente (b_j, a_j, c_j)

$$\begin{bmatrix} b_{j+1} \\ a_{j+1} \\ c_{j+1} \end{bmatrix} = \begin{bmatrix} b_j \\ a_j \\ c_j \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 g}{\partial b_j^2} & \frac{\partial^2 g}{\partial b_j \partial a_j} & \frac{\partial^2 g}{\partial b_j \partial c_j} \\ \frac{\partial^2 g}{\partial a_j \partial b_j} & \frac{\partial^2 g}{\partial a_j^2} & \frac{\partial^2 g}{\partial a_j \partial c_j} \\ \frac{\partial^2 g}{\partial c_j \partial b_j} & \frac{\partial^2 g}{\partial c_j \partial a_j} & \frac{\partial^2 g}{\partial c_j^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial g}{\partial b_j} \\ \frac{\partial g}{\partial a_j} \\ \frac{\partial g}{\partial c_j} \end{bmatrix}$$

Exemplo

- O Teste Denver de Triagem de Desenvolvimento (DDST) é um teste para triagem de problemas cognitivos e comportamentais em crianças pré-escolares.
- A escala reflete o percentual de crianças de uma determinada faixa etária é capaz de executar uma determinada tarefa.
- As tarefas são agrupadas em quatro categorias:
 - contato social
 - habilidade motora fina
 - linguagem
 - habilidade motora grossa
- Exemplos:
 - sorri espontaneamente (realizada por 90% aos três meses de idade),
 - bate dois blocos uns contra os outros (90% aos 13 meses de idade),
 - fala três palavras que não “mãe” e “pai” (90% aos 21 meses de idade),
 - saltos sobre uma perna (90% de crianças com 5 anos de idade).

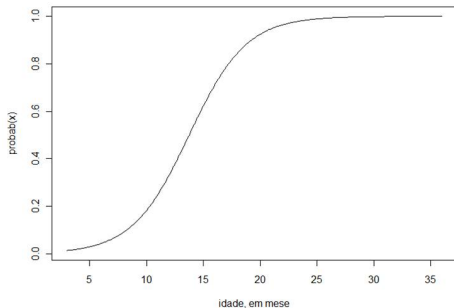
Como o teste é usado?

- Detecção precoce de distúrbios no desenvolvimento é difícil com base em exames clínicos. A ideia então é padronizar as tarefas que crianças deveriam executar numa certa idade.
- Crianças realizam as tarefas-itens apropriadas para sua faixa de idade.
- Considere uma tarefa em que grande maioria (por exemplo, 99%) das crianças naquela faixa de idade consegue executar com sucesso
- Aquelas crianças nesta faixa que NÃO executam estas tarefas são potencialmente problemáticas.

Especificando o teste para um item

- Supõe-se a existência de uma função $f(x)$ que fornece a probabilidade de uma criança com idade x executar o teste.
- As características de $f(x)$:
 - $f(x) \approx 0$ se x é muito baixa
 - $f(x) \approx 1$ se x é muito alta
 - $f(x)$ varia suavemente
- Uma função com estas propriedades é a função logística, mostrada a seguir.

Função logística



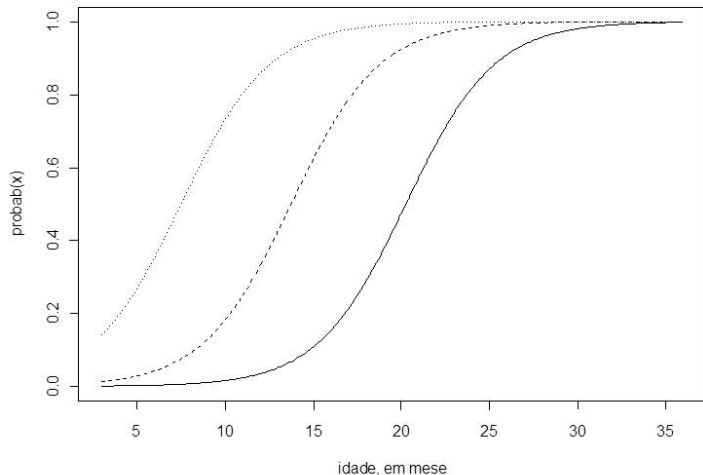
$$p(x) = \frac{1}{1 + \exp\{-\alpha - \beta(x - x_m)\}}$$

Interpretação

- x_m = idade de referência.
- $p(x_M) = \frac{1}{1 + \exp\{-\alpha\}}$.
- Então, α está associado com a probabilidade de executar a tarefa à idade x_m .
- β é a velocidade do crescimento na idade x_m .

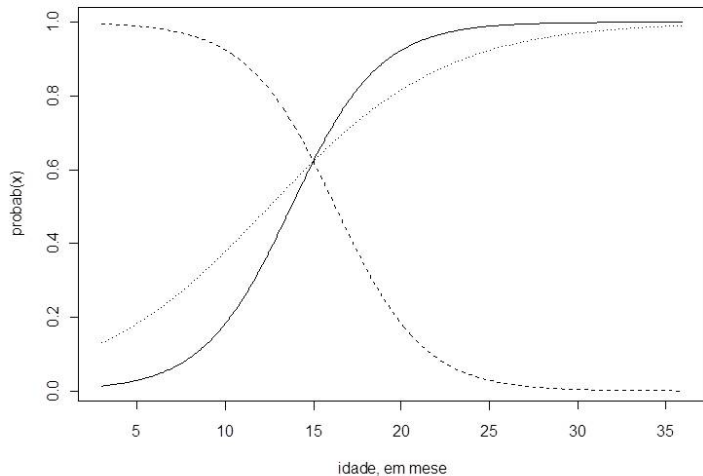
Diferentes parâmetros

Mudando α



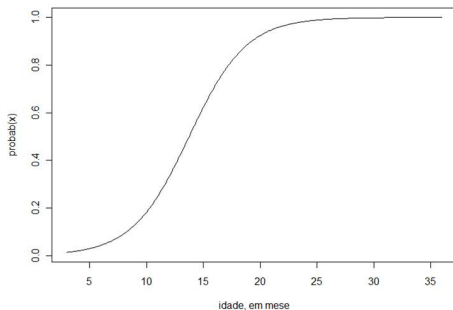
Diferentes parâmetros

Mudando β



Na prática

Se selecionarmos os parâmetros α e β adequadamente, teremos uma curva fixa que descreve como a probabilidade de executar a tarefa evolui com a idade x .



- Como se usa a curva no consultório do pediatra?
- Uma criança com idade x chega.
- Se não executar a tarefa, olha-se $p(x)$.
- Se $p(x) > 0.99$ (digamos), a criança fica sob observação intensiva, faz testes mais caros, etc.
- Tudo depende de ter a curva para cada item.

Fixando a curva para UM ITEM

- Toma-se uma grande amostra de crianças SAUDÁVEIS de idades variadas e pede-se que elas executem a tarefa-item de interesse.
- Sabe-se que são saudáveis por meio de observações intensivas, prolongadas e testes muito caros.
- Verifica-se para cada criança se ela executa ($y = 1$) ou não executa ($y = 0$) a tarefa.
- Se $p(x) \approx 0$ espera-se $y = 0$ e se $p(x) \approx 1$ espera-se $y = 1$.
- Pode-se fazer um gráfico de y versus a idade para as crianças da amostra.

O resultado da amostra

rai-20.jpg

- O problema é determinar a MELHOR curva logística compatível com estes dados.
- Podíamos ajustar no olhometro.
- Mas são muitos itens-tarefas. E queremos um modo mais objetivo de escolha.
- Existe um critério Estatístico de escolha da melhor curva: maximizar a função de log-verossimilhança.

Log-verossimilhança

- As crianças da amostra são rotuladas por $i = 1, \dots, n$.
- Sejam
 - y_i = resultado binário do teste da criança i
 - x_i = idade (em meses) da criança $i - 15$
- A curva logística depende de duas constantes desconhecidas que devem ser estimadas a partir dos dados: $\theta = (\alpha, \beta)$.
- Estas constantes são estimadas obtendo o máximo de uma função chamada de *log-verossimilhança*.
- A função de log-verossimilhança neste problema é dada por

$$L(\alpha, \beta) = \alpha \sum_{i=1}^n y_i + \beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n \log(1 - p(x_i))$$

- onde $p(x_i)$ é função não-linear de α e β .

Maximizando $L(\alpha, \beta)$

- Usamos Newton-Raphson para maximizar $L(\alpha, \beta)$.
- Precisamos do:
 - vetor 2×1 de derivadas parciais de primeira ordem
 - matriz 2×2 de derivadas parciais de segunda ordem
- Temos

$$\begin{aligned} DL(\theta) &= \begin{pmatrix} \frac{\partial \log L}{\partial \alpha} \\ \frac{\partial \log L}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \alpha} \left(\left(\alpha \sum_{i=1}^n y_i + \beta \sum_{i=1}^n x_i y_i \right) + \sum_{i=1}^n \log \left(1 + e^{(\alpha + \beta x_i)} \right) \right) \\ \frac{\partial}{\partial \beta} \left(\left(\alpha \sum_{i=1}^n y_i + \beta \sum_{i=1}^n x_i y_i \right) + \sum_{i=1}^n \log \left(1 + e^{(\alpha + \beta x_i)} \right) \right) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n y_i - \left(\sum_{i=1}^n \frac{e^{(\alpha + \beta x_i)}}{1 + e^{(\alpha + \beta x_i)}} \right) \\ \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n \frac{x_i e^{(\alpha + \beta x_i)}}{1 + e^{(\alpha + \beta x_i)}} \right) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i - \sum_{i=1}^n p_i \\ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i p_i \end{pmatrix} \end{aligned}$$

Matriz de derivadas de 2ª ordem

$$\frac{\partial^2 L(\alpha, \beta)}{\partial(\alpha, \beta)^2} = \begin{bmatrix} -\sum_{i=1}^n p_i(1-p_i) & -\sum_{i=1}^n x_i p_i(1-p_i) \\ -\sum_{i=1}^n x_i p_i(1-p_i) & -\sum_{i=1}^n x_i^2 p_i(1-p_i) \end{bmatrix}.$$

Equação de iteração

A equação de iteração de Newton-Raphson fica:

$$\begin{aligned} \begin{bmatrix} \alpha_{(m+1)} \\ \beta_{(m+1)} \end{bmatrix} &= \begin{bmatrix} \alpha_{(m)} \\ \beta_{(m)} \end{bmatrix} - J_n^{-1} \nabla L_n \\ &= \begin{bmatrix} \alpha_{(m)} \\ \beta_{(m)} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n p_i^{(m)}(1 - p_i^{(m)}) & \sum_{i=1}^n x_i p_i^{(m)}(1 - p_i^{(m)}) \\ \sum_{i=1}^n x_i p_i^{(m)}(1 - p_i^{(m)}) & \sum_{i=1}^n x_i^2 p_i^{(m)}(1 - p_i^{(m)}) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n [y_i - p_i^{(m)}] \\ \sum_{i=1}^n x_i [y_i - p_i^{(m)}] \end{bmatrix}. \end{aligned}$$

onde os valores $p_i = p(x_i) = 1/(1 + \exp(-(\alpha + \beta x_i)))$ são calculados com o valor corrente $(\alpha_{(m)}, \beta_{(m)})$ dos parâmetros.

Ver código em Scilab.

Exemplo

- Foi feito um estudo estatístico para determinar o efeito de alguns fatores no tempo de sobrevida de pacientes diagnosticados com um certo tipo de câncer fatal.
- Vamos considerar aqui apenas os fatores: sexo e estágio do câncer.
- Sabe-se que a distribuição de probabilidade do tempo de sobrevida depende do estágio em que o câncer foi diagnosticado e do sexo do indivíduo.
- Para cada paciente diagnosticado com este câncer, temos as seguintes informações:
 - Tempo de sobrevida (em meses).
 - Sexo: 0 se homem e 1 se mulher.
 - Estágio do câncer: um valor entre 0 a 100 que mede o nível de progressão do câncer no momento do diagnostico, um estágio mais avançado correspondendo a valores maiores.

Exemplo: câncer

- Objetivo é determinar os parâmetros da distribuição de probabilidade de forma que, para um paciente futuro com este câncer, possamos calcular probabilidades tais como:
 - Dado que o paciente é um homem e possui estágio = 35, qual é a probabilidade de sobreviver pelo menos dois anos mais?
- O estudo levou em conta mais fatores além de sexo e estágio mas vamos ignorar estes aspectos.
- No final, obter as probabilidades reduz-se a um problema de maximização de uma função multivariada.

Exemplo: câncer

- Os indivíduos do estudo são rotulados por $i = 1, \dots, n$
- Sejam
 - y_i = tempo de sobrevivência (em meses) do paciente i
 - $x_{1i} = 0$ se o i -ésimo paciente é mulher e $x_{1i} = 1$ se homem
 - x_{2i} é o estágio do câncer do i -ésimo paciente
- O cálculo de probabilidades depende de três constantes desconhecidas que devem ser estimadas a partir dos dados:
 $\theta = (\beta_0, \beta_1, \beta_2)$
- Estas constantes são estimadas obtendo o máximo de uma função chamada de log-verossimilhança.

Os dados de 12 pacientes

i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	3.19	16.87	24.65	2.04	5.73	1.03	6.02	42.41	36.08	7.34	24.88	5.90
x_{1i}	0	0	0	0	0	1	1	1	1	1	1	1
x_{2i}	11	67	92	32	85	36	20	69	58	47	100	72

- Seja $\bar{x}_1 = (0 + \dots + 1)/12 = 0.17$, a media aritmética dos valores x_{1i} , média tirada sobre o conjunto de 12 pacientes do estudo.
- Seja também $\bar{x}_2 = (11 + 67 + \dots + 72)/12 = 53.25$.

Exemplo: câncer

- O objetivo é maximizar em $\theta = (\beta_0, \beta_1, \beta_2)$ a função de log-verossimilhança dada por

$$\ell(\theta) = n(\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2) - \sum_{i=1}^n y_i e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}$$

- Onde $n = 12$ no nosso estudo (número de pacientes).
- Usamos Newton-Raphson multivariado.
- Para montar a equação de iteração precisamos do **VETOR** 3×1 de derivadas parciais de primeira ordem e da **MATRIZ** 3×3 de derivadas parciais de segunda ordem.

Equação de iteração

- Defina $\lambda_i(\theta) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$.
- O vetor de derivadas de primeira ordem:

$$D\ell(\theta) = \begin{bmatrix} \partial\ell/\partial\beta_0 \\ \partial\ell/\partial\beta_1 \\ \partial\ell/\partial\beta_2 \end{bmatrix} = \begin{bmatrix} -n + \sum_i y_i \lambda_i(\theta) \\ -n\bar{x}_1 + \sum_i y_i x_{1i} \lambda_i(\theta) \\ -n\bar{x}_2 + \sum_i y_i x_{2i} \lambda_i(\theta) \end{bmatrix}$$

- A matriz de derivadas de segunda ordem:

$$\begin{aligned} D^2\ell &= \begin{bmatrix} \partial^2\ell/\partial\beta_0^2 & \partial^2\ell/\partial\beta_1\partial\beta_0 & \partial^2\ell/\partial\beta_2\partial\beta_0 \\ \partial^2\ell/\partial\beta_0\partial\beta_1 & \partial^2\ell/\partial\beta_1^2 & \partial^2\ell/\partial\beta_2\partial\beta_1 \\ \partial^2\ell/\partial\beta_0\partial\beta_2 & \partial^2\ell/\partial\beta_1\partial\beta_2 & \partial^2\ell/\partial\beta_2^2 \end{bmatrix} \\ &= - \begin{bmatrix} \sum_i y_i \lambda_i(\theta) & \sum_i y_i x_{1i} \lambda_i(\theta) & \sum_i y_i x_{2i} \lambda_i(\theta) \\ \sum_i y_i x_{1i} \lambda_i(\theta) & \sum_i y_i x_{1i}^2 \lambda_i(\theta) & \sum_i y_i x_{1i} x_{2i} \lambda_i(\theta) \\ \sum_i y_i x_{2i} \lambda_i(\theta) & \sum_i y_i x_{1i} x_{2i} \lambda_i(\theta) & \sum_i y_i x_{2i}^2 \lambda_i(\theta) \end{bmatrix} \end{aligned}$$

Exemplo: câncer

- Vamos começar com o valor inicial

$$\theta^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)} \right) = (0, 0, 0)$$

- Ver código Scilab.