Cálculo Diferencial e Integral

Prof. Wagner Hugo Bonat

Curso de Especialização em DATA SCIENCE & BIG DATA

UFPR

Motivação

Problemas convencionais em ciência de dados

- Previsão ou predição $\rightarrow$ O que vai acontecer?
- Classificação $\rightarrow$ Qual o tipo de um determinado objeto?
- Agrupamento $\rightarrow$ Qual a melhor forma de agregar objetos?
- Prescrição $\rightarrow$ O que devo fazer?

Problemas convencionais em ciência de dados

- Como resolvê-los?
  - Em geral usamos algum tipo de **modelo.**
- O que é um modelo?
  - Representação simplificada da realidade.
- Qual o objetivo de um modelo?
  - Representar como o cientista imagina ou supõe que a realidade está sendo gerada e refletida por meio dos dados.

- Características de um bom modelo
  - Deve representar os principais aspectos do fenômeno sendo avaliado.
  - Pode conter uma ou mais quantidades desconhecidas (parâmetros).
  - Deve permitir generalizações.
  - Deve fornecer um resumo rápido e interpretável do fenômeno em estudo.
  - Deve ser matematicamente preciso e coerente.

Funções

Funções

- Quanto devo pagar pelo seguro do meu carro?
  - Bem, aí depende ...
  - Qual é o seu carro? De qual ano? Que modelo?
  - Qual o uso? Onde você estaciona?
  - Qual a sua idade?
  - Quanto tempo faz que você dirige?
  - E muito mais!!
- Situação complexa, vamos simplificar
  - custo = f(suas características).
- Ilustração do fluxo

Funções

- **Definição:** uma **função** escrita como $y = f(x)$ associa um número y a cada valor de x.
- x é chamada de variável **independente**.
- **Domínio de** $f(x)$ é a faixa de valores que x pode assumir.
- y é chamada de variável **dependente**.
- **Imagem de** $f(x)$ é a faixa de valores que y pode assumir.
- Resumindo temos,

$$\xrightarrow[\text{Independente}]{x \in D} f(x) \xrightarrow[\text{Dependente}]{y \in I}$$

Funções

- O **domínio** e **imagem** de uma função são intervalos.
- Tipos de intervalos:
  - Intervalo aberto **não contém** as extremidades: Notação (a,b).
  - Intervalo fechado **contém** as extremidades: Notação $[a,b]$.
- O que entra e o que sai de uma função?
  - **Naturais:** $\mathbb{N} = \{0,1,2,3,\ldots\}$.
  - **Inteiros:** $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1,2,3, \ldots\}$
  - **Racionais** $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\}$.
  - **Irracionais:** Conjunto de números que não são racionais.
  - **Reais:** União de todos os números mencionados acima, notação $\mathfrak{R}$.
- Distinção importante $\mathfrak{R}$ (`double`) e $\mathbb{Z}$ (`integer`).

Computacionalmente

- Considere a função $y = x^2$.
- Em R temos

```
minha_funcao <- function(x) {
  y <- x^2
  return(y)
}
```

- Avaliando a função em alguns pontos.

```
x_vec <- c(-5, -4, -3, -2, -1,
           0, 1, 2, 3, 4, 5)
minha_funcao(x = x_vec)
```

```
##  [1] 25 16  9  4  1  0  1  4  9 16 25
```

Funções unidimensionais

- Uma função $y = f(x)$ é dita ser de apenas uma variável (unidimensional).
- Pode ser desenhada em um espaço bidimensional, o chamado $\Re^2$.
- O espaço $\Re^2$ é formado por todas as duplas ordenadas de valores reais.
- A variável dependente y é representada no eixo vertical.
- A variável dependente x é representada no eixo horizontal.

Computacionalmente

- Desenho do gráfico da função

```r
## Avaliando a função
y <- minha_funcao(x = x_vec)

## Gráfico da função
plot(y ~ x_vec, xlab = "x", type = "l",
     ylab = expression(y = f(x)))
points(x_vec,y)
```

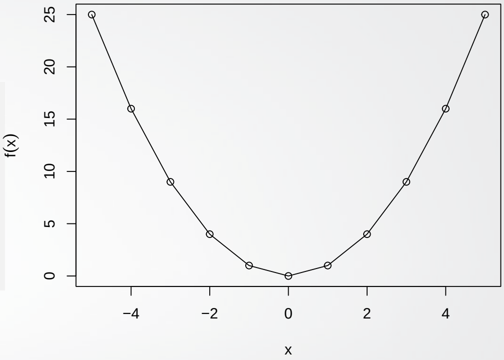


Figura 1. Exemplo de gráfico de função unindo os pontos de avaliação.

Funções parametrizadas

- **Definição – parâmetro** é uma quantidade conhecida que indexa ou parametriza uma determinada função.
- Os parâmetros mudam o comportamento da função e descrevem quantidades/características de interesse.
- Notação: $y = f(x; \theta)$, onde θ denota o parâmetro.
- O conjunto de valores que θ pode assumir é chamado de espaço paramétrico.
- Notação $\theta \in \Theta$.
- Exemplo: $y = (x - \theta)^2$
- Computacionalmente

```
fx = function(x, theta) {
  out <- (x - theta)^2
  return(out)
}
```

Desenho do gráfico da função

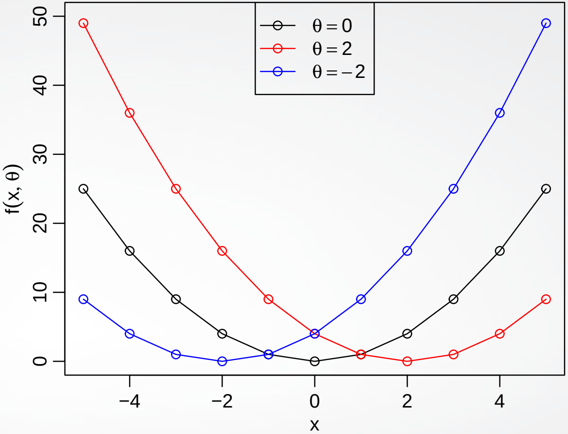


Figura 2. Exemplo de função parametrizada.

Funções com vários parâmetros

- Em geral uma função pode ter vários parâmetros.
- O ideal é que cada parâmetro controle um aspecto da função.
- Exemplo: $y = f(x; \theta)$, onde θ é um vetor de parâmetros.
- Função com dois parâmetros:

$$y = \frac{(x - \theta_1)^2}{\theta_2}.$$

Desenho do gráfico da função

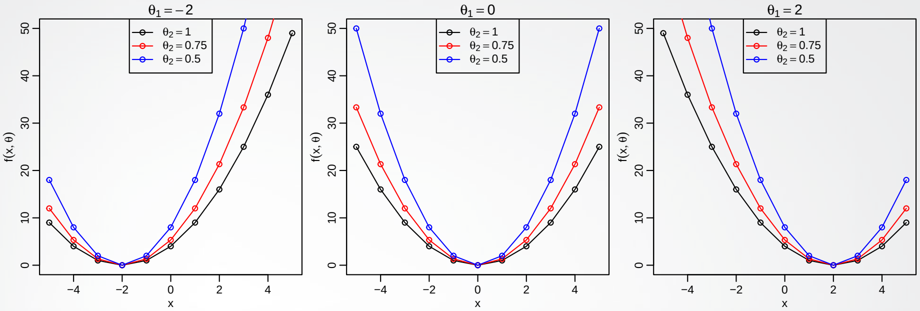


Figura 3. Função parametrizada.

Declividade

- A **declividade** mede a variação Δ no valor de y dividido pela variação no valor de x, ou seja, declividade é $\frac{\Delta y}{\Delta x}$.
- A declividade do desenho de uma função pode ser constante (A), positiva (B) ou negativa (C).

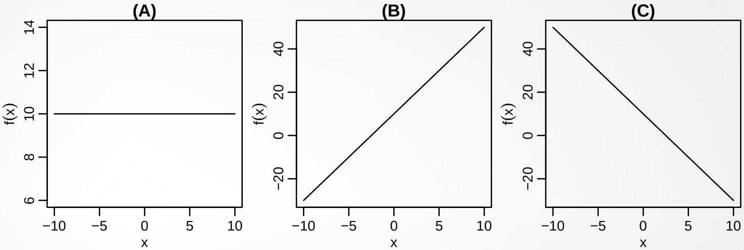


Figura 4. Exemplos de declividade.

- O **intercepto vertical** é o ponto no qual o gráfico cruza o eixo vertical e é obtido quando $x = 0$.

Funções com duas ou mais variáveis independentes

Funções com duas ou mais variáveis independentes

- **Definição** - uma **função** escrita como $y = f(x)$ associa um número y a cada vetor de entrada x.
- $x = (x_1, \dots, x_p)^\top$ denota um vetor linha transposto (vetor coluna).
- Exemplo: considere a função de duas variáveis x_1 e x_2 definida por

$$f(x_1, x_2) = \sqrt{25 - x_1^2 - x_2^2},$$

avalie a função nos pontos $x = (0, 0)$, $x = (3, 0)$ e desenhe seu gráfico.

- Avaliando nos pontos

$$y = \sqrt{25 - 0^2 - 0^2} = 5 \quad \text{e} \quad y = \sqrt{25 - 3^2 - 0^2} = 4.$$

Computacionalmente

- Implementação computacional

```r
fx1x2 <- function(x) {
  y = sqrt(25 - x[1]^2 - x[2]^2)
  return(y)
}

entrada1 <- c(0, 0)
entrada2 <- c(3, 0)
fx1x2(x = entrada1)
```

```
## [1] 5
```

```r
fx1x2(x = entrada2)
```

```
## [1] 4
```

- Avaliando uma função bidimensional.

```r
entrada <- matrix(c(entrada1, entrada2),
                  ncol = 2, nrow = 2,
                  byrow = TRUE)
entrada
```

```
##      [,1] [,2]
## [1,]    0    0
## [2,]    3    0
```

```r
saida <- c()
for(i in 1:2) {
  saida[i] <- fx1x2(entrada[i,])
}
saida
```

```
## [1] 5 4
```

- O gráfico da função é o conjunto das triplas ordenadas (y, x_1, x_2) que satisfazem a função.

Passo-a-passo para desenhar funções bidimensionais

- Neste caso estamos no espaço $\Re^3$.
- (A) Montar uma grade de valores combinando valores para x_1 com valores para x_2.
- (B) Avaliar a função em cada um dos pontos criados.
- (C) Representar o valor da função no gráfico. Neste caso usando uma paleta de cores.

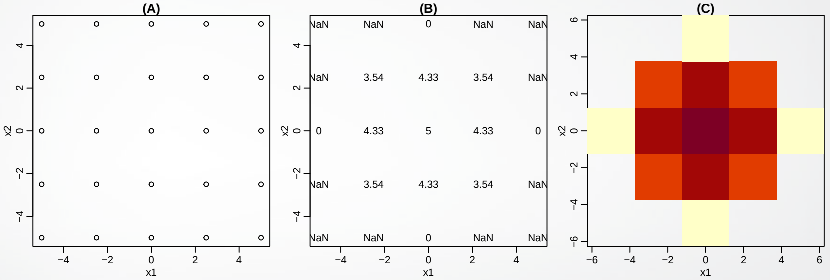


Figura 5. Passo-a-passo para desenhar uma função de duas variáveis independentes.

Gráficos bidimensionais

- Em geral usamos uma grade mais precisa.

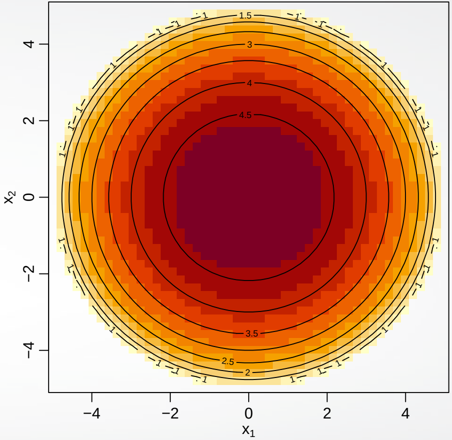


Figura 6. Ilustração do gráfico de uma função de duas variáveis de entrada.

Funções multidimensionais

- **Definição** – uma **função** escrita como $y = f(x; \theta)$ associa um número y a cada vetor de entrada x e θ denota um vetor de parâmetros conhecidos.
- Para funções com mais de duas variáveis de entrada não temos uma forma simples de representação gráfica.
- Em termos práticos as funções vão representar ou modelar situações reais.
- Precisamos de funções flexíveis para representar fenômenos complexos.

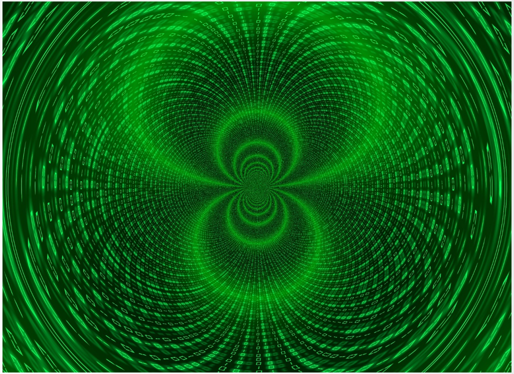


Figura 7. Image by Darwin Laganzon from Pixabay.

Funções especiais

Funções polinômiais

- Funções polinômiais são funções do tipo $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots \beta_p x^p$.
- Exemplo: **funções polinômiais** de grau até três.
  - Função linear: $y = \beta_0 + \beta_1 x$.
  - Função quadrática: $y = \beta_0 + \beta_1 x + \beta_2 x^2$.
  - Função cúbica: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.
- O gráfico de uma função quadrática é uma parábola aberta para cima se $\beta_2 > 0$ ou para baixo se $\beta_2 < 0$.

- Graficamente

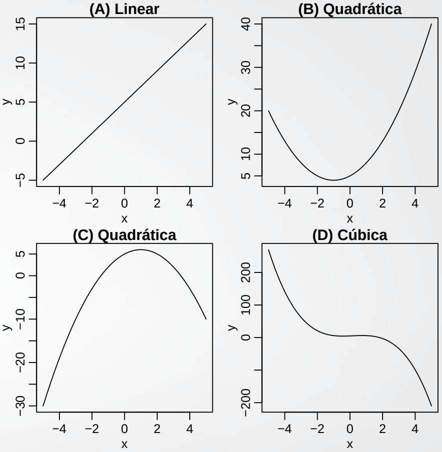


Figura 8. Exemplos de gráficos de funções polinômiais.

Funções do tipo potência

- Funções do **tipo potência** são funções da forma
$$y = x^a,$$
em que a é um expoente constante.
- Por definição, $x^0 = 1$ e note que um número sem expoente está elevado a 1.

- Propriedades importantes:

1. $x^a(x^c) = x^{a+c}$;
2. $(x^a)^c = x^{ac}$;
3. $(xz)^a = x^a(z^a)$;
4. $\left(\frac{x}{z}\right)^c = \frac{x^c}{z^c}$;
5. $\frac{1}{x^a} = x^{-a}$;
6. $\frac{x^a}{x^c} = x^{a-c}$;
7. $\sqrt{x} = x^{1/2}$;
8. $\sqrt[a]{x} = x^{1/a}$;
9. $\sqrt[c]{x^a} = x^{a/c}$.

Funções exponenciais

- **Funções exponenciais** são funções do tipo $y = a^x$ onde a é maior que zero e diferente de 1 e x é o expoente.
- **Funções exponenciais naturais** são funções exponenciais que tem como base $e = \lim_{n \to \infty} [1 + (1/n)]^n = 2.718281828$.
- Propriedades importantes:

1. $e^0 = 1$.
2. $e^1 = e = 2.71828$.
3. $e^a(e^b) = e^{a+b}$.
4. $(e^a)^b = e^{ab}$.
5. $\frac{e^a}{e^b} = e^{a-b}$.

Funções logarítmicas

- **Funções logarítmicas** ou logaritmo é a potência à qual uma dada base deve ser elevada para se obter um particular número.
- Logaritmos comuns utilizam a base 10 e são escritos $\log_{10}$.
- Por exemplo, uma vez que $10^2 = 100$, 2 é o log de 100.
- Para qualquer função exponencial $y = a^x$, onde a é a base e x o expoente, $\log_a y = x$ x é a potência à qual a deve ser elevado, para obter-se y.

Relações entre funções logarítmicas e exponenciais.

- Se $\log_{10} y = 2x$, então $y = 10^{2x}$.
- Se $\log_a y = xz$, então $y = a^{xz}$.
- Se $\ln y = 5t$, então $y = e^{5t}$.
- Se $y = a^{3x}$, então $\log_a y = 3x$.
- Se $y = 10^{6x}$, então $\log_{10} y = 6x$.
- Se $y = e^{t+1}$, então $\ln y = t + 1$.

Outras funções de interesse

- Sigmóide ou logística: $y = \frac{1}{1+e^{-x}}$.
- Tangente hiperbólica: $y = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$.
- Linear retificada (ReLU): $y = \max\{0,x\}$.
- *Leaky ReLU*: $y = \max\{\alpha x, x\}$, onde α é uma parâmetro conhecido.

Desenho do gráfico das funções

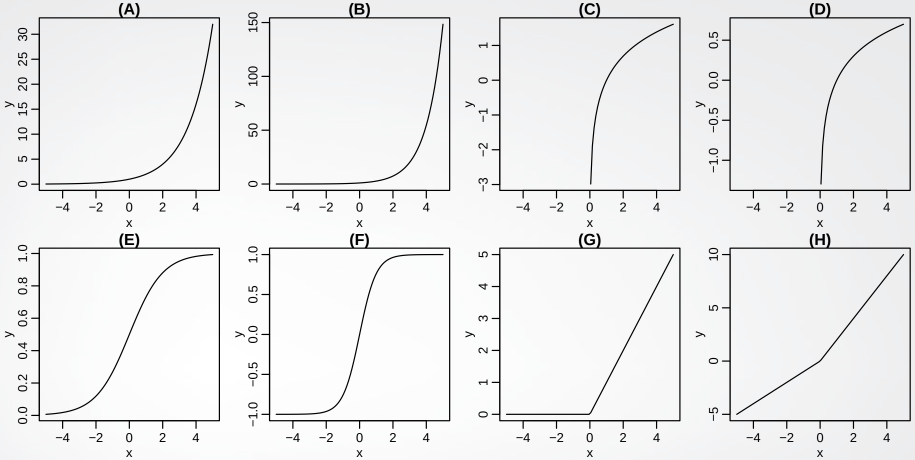


Figura 9. Exemplos de gráficos de funções: (A) potência (a = 2); (B) exponencial natural; (C) Log natural; (D) Log 10; (E) Sigmóide; (F) Tangente hiperbólica; (G) ReLU; (H) Leaky ReLU.

Limites e continuidade

Limite de uma função

- **Definição** - se uma função $f(x)$ se aproxima de um número L conforme x tende a um número a vindo da direita ou da esquerda, dizemos que o limite de $f(x)$ tende a L quando x tende a a.
- Notação
$$\lim_{x \to a} f(x) = f(a) = L.$$
- O limite pode não existir.
- Se o limite de uma função existe ele é único.
- Considere o limite
$$\lim_{x \to 1} (x + 1) = 2.$$

Exemplo

- Considere o limite

$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = ?$$

- Computacionalmente

```r
fx <- function(x) {
  out <- (x^2 - 1)/(x - 1)
  return(out)
}
fx(x = 1)
```

```
## [1] NaN
```

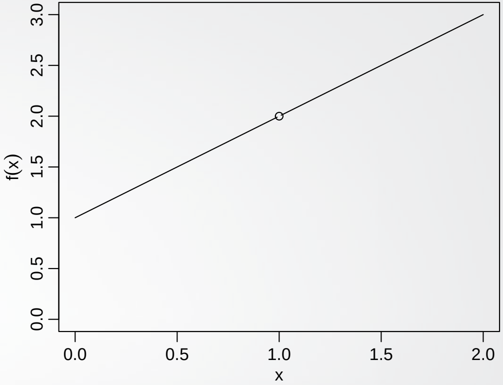


Figura 10. Desenho do gráfico da função.

Exemplo

- Note que

$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = \lim_{x \to 1} \frac{(x+1)(x-1)}{x - 1} = \lim_{x \to 1} (x + 1) = 2.$$

- **Definição intuitiva:** o limite de uma função é o valor que achamos natural para ela em um determinado ponto.

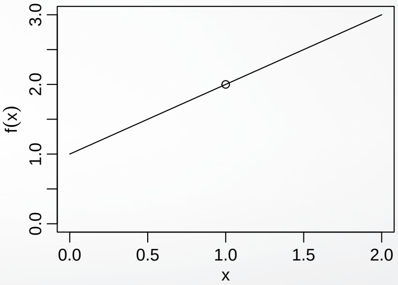


Figura 11. Desenho do gráfico da função.

Continuidade de uma função

- **Definição** – dizemos que uma função é **contínua** em $x = a$ se três condições forem satisfeitas:
  - $f(a)$ existe,
  - $\lim_{x\to a} f(x)$ existe e
  - $\lim_{x\to a} f(x) = f(a)$.
- Continuidade significa que pequenas variações na variável independente levam a pequenas variações na variável dependente.
- **Teorema do valor intermediário:** se a função $f(x)$ é contínua no intervalo fechado $[a,b]$, então existe pelo menos um número c em $[a,b]$ tal que $f(c) = M$.
- Implicação: se $f(x)$ é contínua seu gráfico não contém salto vertical.
- Em geral podemos pensar em funções contínuas como sendo funções suaves.

Função não contínua

- Considere a função não continua em 0.

$$\lim_{x \to 0} \frac{|x|}{x} = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases}$$

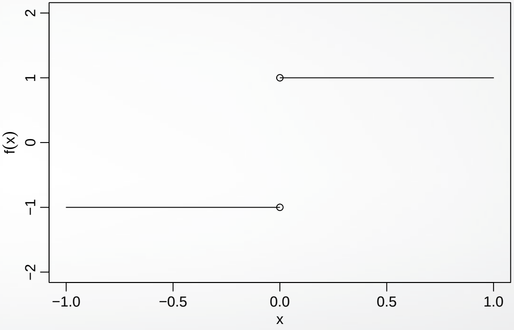


Figura 12. Função descontinua.

Propriedades de limites

Se $\lim_{x \to p} f(x) = L_1$ e $\lim_{x \to p} g(x) = L_2$, então

- $\lim_{x \to p} [f(x) + g(x)] = \lim_{x \to p} f(x) + \lim_{x \to p} g(x) = L_1 + L_2.$
- $\lim_{x \to p} kf(x) = k \lim_{x \to p} f(x) = kL_1.$
- $\lim_{x \to p} f(x)g(x) = \lim_{x \to p} f(x) \lim_{x \to p} g(x) = L_1 L_2.$
- $\lim_{x \to p} \frac{f(x)}{g(x)} = \frac{L_1}{L_2}$, desde que $L_2 \neq 0$.

Aquecimento $\longrightarrow$ DONE!

Figura 13. Image by Simona Robová from Pixabay.

Derivadas

Definição

- **Definição** - derivada ordinária, derivada primeira, ou simplesmente, derivada de uma função $y = f(x)$ em um ponto $x = a$ no domínio de f é representada por $\frac{dy}{dx}$, y', $\frac{df}{dx}$ ou $f'(a)$ é o valor
$$\frac{dy}{dx}|_{x=a} = f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$
- Interpretação da derivada
- Taxa de mudança instântanea.
- No limite quando $x \to a$ a derivada é a reta tangente ao ponto $(a, f(a))$.
- Equação da reta tangente ao ponto a: $y - f(a) = f'(a)(x - a)$.

Exemplo

- Obtenha a derivada de $f(x) = -x^2$.

$$
\begin{aligned}
f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h \to 0} \frac{-(x+h)^2 - (-x^2)}{h} \\
&= \lim_{h \to 0} \frac{-(x^2 + 2xh + h^2) + x^2}{h} \\
&= \lim_{h \to 0} \frac{-x^2 - 2xh - h^2 + x^2}{h} = \frac{-2xh - h^2}{h} \\
&= \lim_{h \to 0} -2x - h = -2x \\
f'(x) &= -2x. \qquad (1)
\end{aligned}
$$

Exemplo

- Obtenha a reta tangente a $f(x)$ nos pontos $x = 2$ e $x = -2$.
- Temos $f(x = 2) = -4$ e $f'(x = 2) = -4$, assim

$$
\begin{aligned}
y - f(x = 2) &= f'(x = 2)(x - 2) \\
y - (-4) &= -4(x - 2) \\
y + 4 = &= -4x + 8 \\
y &= 4 - 4x
\end{aligned}
$$

- Computacionalmente
- $f(x)$ e $f'(x)$.

```r
fx <- function(x) {
  out <- - x^2
  return(out)
}

f_prime <- function(x) {
  out <- -2*x
  return(out)
}
```

- Equação da reta $y = a + b * x$.

```r
intercept = (fx(x = 2) - f_prime(x = 2)*2)
slope <- f_prime(x = 2)
c(intercept, slope)
```

```
## [1]  4 -4
```

Reta tangente a $f(x)$

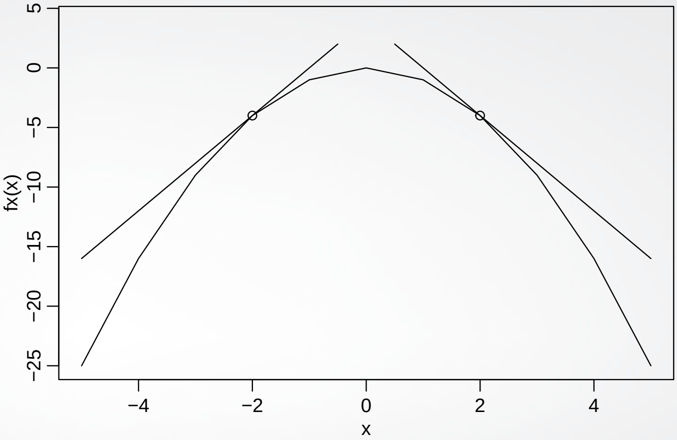


Figura 14. Desenho de uma função e retas tangentes.

Regras de derivação

Regras de derivação

- Seja $n \neq 0$ um natural. São válidas as fórmulas de derivação:
  1. Se $f(x) = c$ então $f'(x) = 0$.
  2. Se $f(x) = x^n$ então $f'(x) = nx^{n-1}$.
  3. Se $f(x) = x^{-n}$ então $f'(x) = -nx^{-n-1}$.
  4. Se $f(x) = x^{1/n}$ então $f'(x) = \frac{1}{n}x^{\frac{1}{n}-1}$.
- Derivada de funções especiais
  5. Se $f(x) = \exp(x)$ então $f'(x) = \exp(x)$.
  6. Se $f(x) = \ln(x)$ então $f'(x) = \frac{1}{x}, x > 0$.
- Sendo, $f(x)$ e $g(x)$ deriváveis em x e c uma constante.
  7. $(f + g)' = f'(x) + g'(x)$.
  8. $(cf)'(x) = cf'(x)$.
  9. $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.
  10. $(\frac{f}{g})'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$.
- Exemplo: obtenha a derivada de $f(x) = 2 + 3x$.
- Solução: $f'(x) = 3$.
- Computacionalmente

```
D(expression(2 + 3*x),
  name = "x")
```

```
## [1] 3
```

Regra da cadeia

- Sejam $y = f(x)$ e $x = g(t)$ duas funções deriváveis, com $I \in D_f$. A função composta $h(t) = f(g(t))$ é derivável, sendo

$$h'(t) = f'(g(t))g'(t), t \in D_g.$$

- Existe uma infinidade de fórmulas de derivação.
- Na prática é comum usar um software de matemática simbólica como o wxMaxima.
- Em `R` as funções `deriv()` e `deriv3()`.

Exemplo regra da cadeia

- Obtenha a derivada de $\sin(2x^3 - 4x)$.

1. Note que temos uma função composta
$$\sin(g(x)), \quad \text{onde} \quad g(x) = 2x^3 - 4x.$$
2. Usando a regra da cadeia temos:
$$f'(g(x)) = \cos(2x^3 - 4x) \quad \text{and} \quad g'(x) = 6x^2 - 4.$$
3. Assim, a derivada fica dada por
$$\cos(2x^3 - 4x) \cdot (6x^2 - 4).$$
4. Computacionalmente

```
D(expression( sin(2*x^3 - 4*x)), name = "x")
```

```
## cos(2 * x^3 - 4 * x) * (2 * (3 * x^2) - 4)
```

Derivadas de ordem superior

Derivadas de ordem superior

- A derivada $f'(x)$ é também chamada de derivada de primeira ordem e mede a variação da função original ou **primitiva**.
- A derivada de segunda ordem denotada por $f''(x)$ mede a taxa de variação da primeira derivada.
- A derivada de terceira ordem $f'''(x)$ mede a taxa de variação da segunda derivada e assim por diante até a $n-$ésima derivada.
- Notação comum: $\frac{d^n y}{dx^n}$ que é interpretada como a $n-$ésima derivada de y em relação a x
- Exemplo: obtenha as derivadas até a ordem 5 da função $y = 2x^4 + 5x^3 + 2x^2$.

$\frac{dy}{dx} = 8x^3 + 15x^2 + 4x$, $\frac{d^2y}{dx^2} = 24x^2 + 30x + 4$, $\frac{d^3y}{dx^3} = 48x + 30$, $\frac{d^4y}{dx^4} = 48$ e $\frac{d^5y}{dx^5} = 0$.

Importância da derivada

Máximos e mínimos

- Dizemos que um ponto c é um **valor máximo relativo** de $f(x)$ se existir um intervalo aberto contendo c, no qual $f(x)$ esteja definida, tal que $f(c) \geq f(x)$ para todo x neste intervalo.
- Dizemos que um ponto c é um **valor mínimo relativo** de $f(x)$ se existir um intervalo aberto contendo c, no qual $f(x)$ esteja definida, tal que $f(c) \leq f(x)$ para todo x neste intervalo.

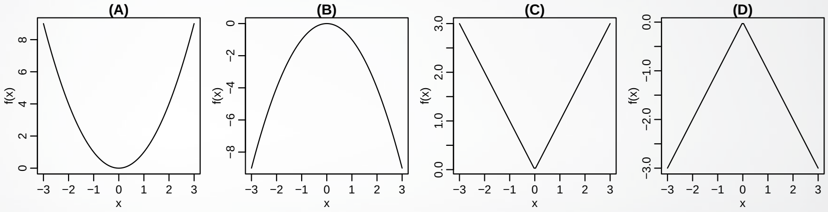


Figura 15. Ilustração de máximo/mínimo relativos.

- Multiplicando a função por -1 invertemos a sua concavidade.

Pontos extremos

- Se $f(x)$ existe para todos os valores de x no intervalo aberto (a,b), e se $f(x)$ tem um extremo relativo em c, em que $a < c < b$, então $f'(c)$ existe e $f'(c) = 0$.
- Implicação – Sendo $f(x)$ diferenciável os pontos extremos de $f(x)$ vão ocorrer quando $f'(x) = 0$.
- $f'(x)$ pode ser igual a zero mesmo não sendo um extremo relativo.

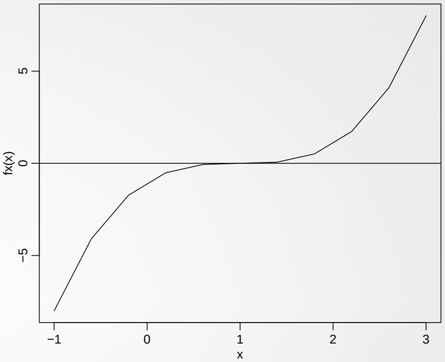


Figura 16. Ilustração de uma função onde derivada zero não é ponto extremo.

Máximos e mínimos

Seja c um ponto extremo de uma função $f(x)$ no qual $f'(c) = 0$, e suponha que $f'(x)$ exista para todos os valores de x em um intervalo aberto contendo c. Se $f''(c)$ existe, então

- Se $f''(c) < 0$, então $f(x)$ tem um **máximo relativo** em c.
- Se $f''(c) > 0$, então $f(x)$ tem um **mínimo relativo** em c.

Concavidade

- Se $f''(c) > 0$ o gráfico de $f(x)$ é côncavo para cima em $(c, f(c))$;
- Se $f''(c) < 0$ o gráfico de $f(x)$ é côncavo para baixo em $(c, f(c))$.

Por que derivadas são importantes?

- Obtenção de máximo ou mínino (relativo).

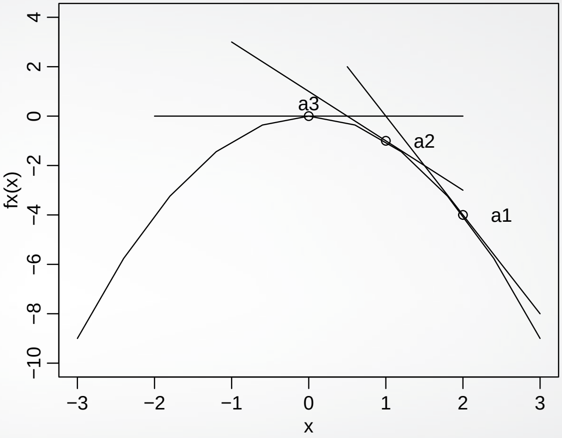


Figura 17. Ilustração de uma função com a reta tangente.

E a ciência de dados??

Figura 18. Image by Ryan McGuire from Pixabay.

Redução de dados

Redução de dados

- Você já trabalha com dados? Se sim,
  - Por qual razão você usa a média ou a mediana como uma medida resumo?
  - Você acha que existe algum procedimento mais geral que leva a obtenção destas medidas resumo?
  - Se sim, como este procedimento está relacionado com o que vimos em relação a funções e seu comportamento?

Figura 19. Image by Peggy und Marco Lachmann-Anke from Pixabay.

Redução de dados

- Suponha que temos um conjunto de observações y_i para $i = 1, \ldots, n$.
- **Objetivo:** resumir a informação contida em y_i em um único número, digamos μ.
- **Problema:** como encontrar μ?
- **Solução:** encontrar o valor μ, tal que $f(\mu) = \sum_{i=1}^{n} (y_i - \mu)^2$, seja a menor possível.
- Uma vez que temos os números observados y_i a única quantidade desconhecida é μ.
- Note que μ é o parâmetro da nossa função.
- A função $f(\mu)$ mede o quanto **perdemos** em representar y_i apenas usando μ.
- Funções perda muito populares são a **perda quadrática**, **perda absoluta**, **minmax** e a **cross entropia**.

Exemplo: redução de dados

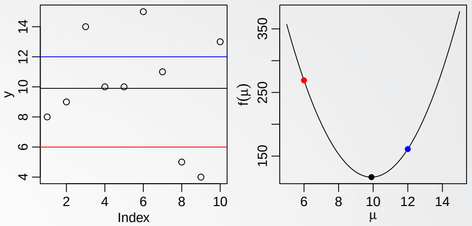
- Funções em R.

```
y <- c(8,9,14,10,10,15,11,5,4,13)

fmu <- function(mu, y) {
  out <- sum((y - mu)^2)
  return(out)
}

fmu <- Vectorize(fmu, "mu")
fmu(mu = c(10, 12), y = y)
```

```
## [1] 117 161
```

```
f_prime <- function(mu, y) {
  out <- -2*sum(y-mu)
  return(out)
}
```

- Graficamente



- Note que o **melhor** resumo dos dados de um número, corresponde ao ponto de mínimo da função $f(\mu) = \sum_{i=1}^{n}(y_i - \mu)^2$.
- Como o mínimo está relacionado com a derivada de $f(\mu)$?

Ponto de mínimo

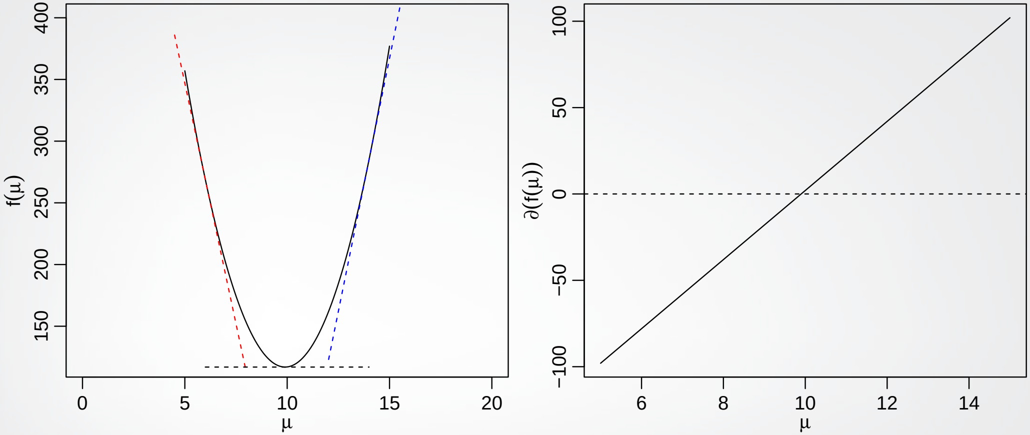


Figura 20. Resumo do processo de minimização da função perda quadrática.

Exemplo: redução de dados

- No ponto de mínimo/máximo a inclinação da reta tangente a $f(\mu)$ é zero.
- Denote por $\hat{\mu}$ o ponto de mínimo/máximo de $f(\mu)$, então $f'(\hat{\mu}) = 0$.
- Assim, temos (regra da cadeia!!)

$$
\begin{aligned}
f'(\mu) &= 2\sum_{i=1}^{n}(y_i - \mu)\frac{d}{d\mu}(y_i - \mu) \\
&= 2\sum_{i=1}^{n}(y_i - \mu)(-1) = -2\sum_{i=1}^{n}(y_i - \mu).
\end{aligned}
$$

Exemplo: redução de dados

- Agora precisamos achar o ponto $\hat{\mu}$ tal que $f'(\hat{\mu}) = 0$.

$$
\begin{aligned}
f'(\hat{\mu}) &= 0 \\
-2\sum_{i=1}^{n}(y_i - \hat{\mu}) &= 0 \\
-\sum_{i=1}^{n} y_i + n\hat{\mu} &= 0 \\
n\hat{\mu} &= \sum_{i=1}^{n} y_i \\
\hat{\mu} &= \frac{\sum_{i=1}^{n} y_i}{n}.
\end{aligned}
$$

Comentários

- Por qual razão você usa a média ou a mediana como uma medida resumo?
  - Minimiza a perda quadrática.
  - Medida ótima no sentido de perda quadrática.
- Você acha que existe algum procedimento mais geral que leva a obtenção destas medidas resumo?
  - Especificação do modelo.
  - Escolha da função perda.
  - Treinamento (otimização).
- Se sim, como este procedimento está relacionado com o que vimos em relação a funções e seu comportamento?
  - Estudar o comportamento de funções.

Derivadas parciais

Derivadas parciais

- Uma função pode ter mais do que uma variável independente.
- A **derivada parcial** mede a taxa de variação instantânea da variável dependente (y) com relação a variável independente x_1, quando a outra variável independente x_2 é mantida constante.
- Como obter a **derivada parcial**?
  - A **derivada parcial** em relação a x_1 é obtida derivando $f(x_1, x_2)$ "fingindo" que x_2 é uma constante.
  - A derivada parcial de $f(x_1, x_2)$ em relação a x_2 é obtida derivando $f(x_1, x_2)$ mantendo x_1 constante.
  - A diferenciação parcial segue as mesmas regras da diferenciação ordinária.

Exemplo

Obtenha as derivadas parciais em relação a x_1 e x_2 de $y = 5x_1^3 + 3x_1x_2 + 4x_2^2$.

$$\frac{\partial y}{\partial x_1} = 15x_1^2 + 3x_2.$$

$$\frac{\partial y}{\partial x_2} = 3x_1 + 8x_2.$$

Derivadas parciais de ordem superior

- **Derivadas parciais de segunda ordem**

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} \quad \text{e} \quad \frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}$$

indica que a função foi diferenciada parcialmente em relação a x_1 ou x_2 duas vezes.

- **Derivada parcial cruzada (ou mista)**

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2}$$

indica que primeiro derivamos em x_1 e depois em x_2.

- A ordem da derivada cruzada não importa (se ambas contínuas), ou seja

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1}.$$

Exemplo: derivadas parciais de segunda ordem

- Obtenha as derivadas parciais de até segunda ordem em relação a x_1 e x_2 de $y = 7x_1^3 + 9x_1x_2 + 2x_2^5$.
- Derivadas parciais de primeira ordem

$$\frac{\partial y}{\partial x_1} = 21x_1^2 + 9x_2, \qquad \frac{\partial y}{\partial x_2} = 9x_1 + 10x_2^4.$$

- Derivadas parciais de segunda ordem (segunda derivadas direta)

$$\frac{\partial^2 y}{\partial x_1^2} = 42x_1, \qquad \frac{\partial^2 y}{\partial x_2^2} = 40x_2^3.$$

- Derivadas parciais de segunda ordem (termos cruzados)

$$\frac{\partial^2 y}{\partial x_1 x_2} = \frac{\partial 21x_1^2 + 9x_2}{\partial x_2} = 9, \qquad \frac{\partial^2 y}{\partial x_2 x_1} = \frac{\partial 9x_1 + 10x_2^4}{\partial x_1} = 9.$$

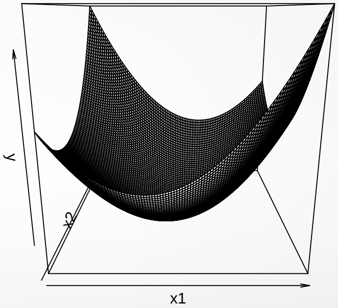
Máximos e mínimos funções muldimensionais

- Pontos críticos: as derivadas parciais de primeira ordem devem ser iguais a zero **simultaneamente**.
- Derivadas parciais de segunda ordem no ponto crítico forem ambas **positivas** $\rightarrow$ **ponto de mínimo.**
- Derivadas parciais de segunda ordem no ponto crítico forem ambas **negativas** $\rightarrow$ **ponto de máximo.**
- Outras situações ver material suplementar.

Exemplo

Considere a função $y = 6x_1^2 - 9x_1 - 3x_1x_2 - 7x_2 + 5x_2^2$. Encontre os pontos críticos e determine se são de máximo ou minímo.

- Graficamente



Exemplo (cont.)

- Calcular as derivadas parciais de primeira ordem da função $y = 6x_1^2 - 9x_1 - 3x_1x_2 - 7x_2 + 5x_2^2$.
- Derivando em x_1, temos

$$\frac{\partial y}{\partial x_1} = 12x_1 - 9 - 3x_2.$$

- Derivando em x_2, temos

$$\frac{\partial y}{\partial x_2} = -3x_1 - 7 + 10x_2.$$

Exemplo (cont.)

- Resolver o sistema de equações

$$\begin{aligned} 12x_1 - 9 - 3x_2 &= 0 \\ -3x_1 - 7 + 10x_2 &= 0. \end{aligned}$$

- Solução: $x_1 = 1$ e $x_2 = 1$.

Exemplo (cont.)

- Verificar se o ponto encontrado é de mínimo calculando a segunda derivada parcial e avaliando o seu sinal.

$$\frac{\partial^2 y}{\partial x_1^2} = \frac{\partial}{\partial x_1}(12x_1 - 9 - 3x_2) = 12, \quad \frac{\partial^2 y}{\partial x_2^2} = \frac{\partial}{\partial x_2}(3x_1 - 7 + 10x_2) = 10.$$

- Calcular as derivadas cruzadas e verificar se o produto das derivadas principais é maior que o produto das cruzadas

$$\frac{\partial^2 y}{\partial x_1 x_2} = \frac{\partial 12x_1 - 9 - 3x_2}{\partial x_2} = -3, \qquad \frac{\partial^2 y}{\partial x_2 x_1} = \frac{\partial - 3x_1 - 7 + 10x_2}{\partial x_2} = -3.$$

Exemplo (cont.)

Assim, temos que

$$
\begin{aligned}
\frac{\partial^2 y}{\partial x_1^2}\frac{\partial^2 y}{\partial x_2^2} &> \left(\frac{\partial^2 y}{\partial x_1 x_2}\right)^2 \\
12 \cdot 10 &> (-3)^2 \\
120 > 9.
\end{aligned}
$$

- A função está em um ponto de mínimo quando examinada de todas as direções.

Gradiente e Hessiano

Gradiente

- Derivadas de primeira e segunda ordem aparecem com tanta frequência que receberam nomes especiais.
- O **vetor gradiente** de uma função $f(x_1,x_2)$ é o vetor composto pelas derivadas primeira de $f(x_1,x_2)$ em relação a x_1 e x_2,

$$\nabla f(x_1,x_2) = \left(\frac{\partial f(x_1,x_2)}{\partial x_1}, \frac{\partial f(x_1,x_2)}{\partial x_2} \right)^\top .$$

- A definição estende-se naturalmente para funções multidimensionais.
- Sendo, $f(x)$ onde x é um vetor $p \times 1$ de variáveis independentes o **vetor gradiente** de $f(x)$ é dado por

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_p} \right)^\top .$$

Hessiano

- A **matriz hessiana** de uma função $f(x_1,x_2)$ é a matriz composta pelas derivadas de segunda ordem de $f(x_1,x_2)$, na seguinte estrutura

$$\mathrm{H} = \begin{pmatrix} \frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} & \frac{\partial f(x_1,x_2)}{\partial x_1 \partial x_2} \\ \frac{\partial f(x_1,x_2)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_2^2} \end{pmatrix}.$$

- E para o caso multidimensional

$$\mathrm{H} = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial f(x)}{\partial x_1 \partial x_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_p \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_p^2} \end{pmatrix}.$$

Séries de Taylor

Séries de Taylor

- Suponha que uma função $f(x)$ é derivável $(n+1)$ vezes em um intervalo contendo $x = x_0$.
- Expansão em **Série de Taylor** de $f(x)$ em torno de $x = x_0$ consiste em reescrever $f(x)$ da seguinte forma:

$$f(x) = f(x_0) + (x - x_0)\frac{df(x)}{dx}|_{x=x_0} + \frac{(x-x_0)^2}{2!}\frac{d^2 f(x)}{dx^2}|_{x=x0} + \quad (2)$$

$$\frac{(x-x_0)^3}{3!}\frac{d^3 f(x)}{dx^3}|_{x=x_0} + \ldots + \frac{(x-x_0)^n}{n!}\frac{d^n f(x)}{dx^n}|_{x=x_0} + R_n(x) \quad (3)$$

onde o termo $R_n(x)$ é chamado de resíduo ou erro, e dado por

$$R_n(x) = \frac{(x-x_0)^{n+1}}{(n+1)!}\frac{d^{n+1} f(x)}{dx^{n+1}}|_{x=\epsilon}$$

sendo ϵ um valor entre x e x_0.

Exemplo

- Seja $f(x) = \exp(x)$. Determine a expansão de Taylor de ordens 1 e 2, de $f(x)$ ao redor de $x_0 = 0$.
- Aproximação de primeira ordem

$$
\begin{aligned}
P_1(x) &= f(x=0) + f'(x=0)(x-0) \\
&= \exp(0) + \exp(0)(x-0) \\
&= 1 + x.
\end{aligned}
$$

- Aproximação de segunda ordem

$$
\begin{aligned}
P_2(x) &= f(x=0) + f'(x=0)(x-0) + \frac{f''(x=0)}{2}(x-0)^2 \\
&= \exp(0) + \exp(0)(x-0) + \frac{\exp(0)}{2}(x-0)^2 \\
&= 1 + x + \frac{1}{2}x^2.
\end{aligned}
$$

Exemplo

- Graficamente

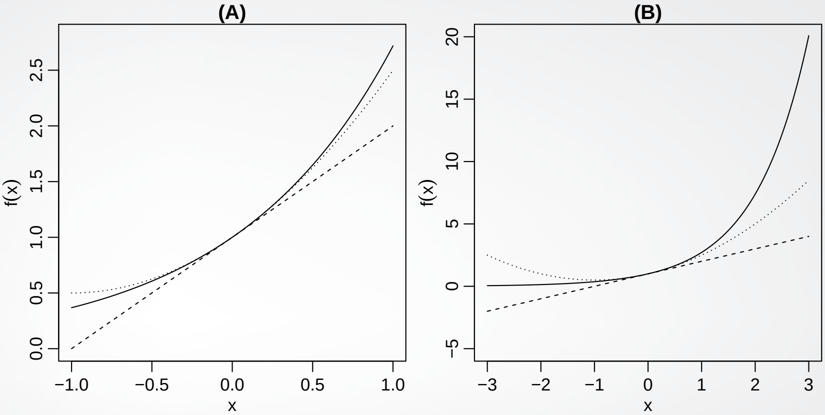
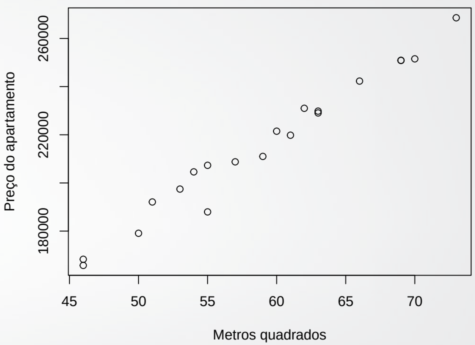


Figura 21. Ilustração da aproximação em Séries de Taylor de primeira e segunda ordem da função exponencial ao redor de zero.

Regressão linear simples

Regressão linear simples

- Regressão linear é uma das técnicas mais populares em ciência de dados.
- **Objetivo:** descrever o comportamento de uma variável dependente y por meio do conhecimento de outra variável independente x.
- Predizer y dado um valor de x.
- Descrever a relação entre y e x.
- Exemplo
  - Como o tamanho (em metros quadrados) de um apartamento está associado ao seu preço (em reais)?
  - Suponha que um conjunto de 20 apartamentos foi medido e avaliado.
- Graficamente



Regressão linear simples

- **Ideia simples!** $\rightarrow$ O preço deve ser uma **função** do tamanho do apartamento.
- Formalização matemática:
  - Denote por y_i para $i = 1, \ldots, n$ o preço do apartamento i e neste caso $n = 20$.
  - Denote por x_i o tamanho do apartamento i em metros quadrados.
- Função relacionando preço $\sim$ tamanho

$$y_i = f^*(x_i).$$

- Qual é a função $f^*(x_i)$?
  - Não conhecemos e em geral nunca vamor conhecer $f^*(x_i)$.
  - Aproximar $f^*(x_i)$ por outra função $f(x_i)$ conhecida.
  - Problema: qual $f(x_i)$ e como fazer a aproximação!

Regressão linear simples

- Uma opção é usar a expansão em série de Taylor para obter uma aproximação.
- Aproximação em série de Taylor de primeira ordem

$$f^*(x) = f^*(x_0) + (x - x_0)f^{*\prime}(x_0) + R_n(x).$$

- Ignorando o termo residual $R_n(x)$

$$f^*(x) \approx f^*(x_0) + (x - x_0)f^{*\prime}(x_0).$$

Regressão linear simples

- Rearranjando os termos obtemos

$$
\begin{aligned}
f^*(x) &\approx \underbrace{\{f^*(x_0) - f^{*\prime}(x_0)x_0\}}_{\beta_0} + \underbrace{f^{*\prime}(x_0)}_{\beta_1}x \\
f^*(x) &\approx \beta_0 + \beta_1 x.
\end{aligned}
$$

- De forma equivalente, temos

$$y_i = \beta_0 + \beta_1 x_i + R_n(x_i),$$

em que o termo $R_n(x_i)$ é o erro cometido em aproximar y_i por $\beta_0 + \beta_1 x_i$.

Regressão linear simples

- Notação usual $\epsilon_i = y_i - (\beta_0 + \beta_1 x_i)$.
- Note que o erro é uma função dos parâmetros desconhecidos β_0 e β_1.
- **Objetivo:** minimizar a soma de quadrados dos erros ou resíduos,

$$
\begin{aligned}
SQ(\beta_0, \beta_1) &= \sum_{i=1}^{n} \epsilon^2(\beta_0, \beta_1)_i \\
SQ(\beta_0, \beta_1) &= \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.
\end{aligned}
$$

Passo-a-passo

- Obter o vetor gradiente

$$\nabla SQ(\beta_0,\beta_1) = \left(\frac{\partial SQ(\beta_0,\beta_1)}{\partial \beta_0}, \frac{\partial SQ(\beta_0,\beta_1)}{\partial \beta_1} \right).$$

- Encontrar $\hat{\beta}_0$ e $\hat{\beta}_1$ tal que

$$\nabla SQ(\hat{\beta}_0,\hat{\beta}_1) = 0.$$

Vetor gradiente

1. Chame $y_i - (\beta_0 + \beta_1 x_i) = \epsilon_i$.
2. Chame $\beta_0 + \beta_1 x_i = \mu_i$.
3. Assim,

$$\nabla SQ(\beta_0,\beta_1) = \left(\frac{\partial SQ(\beta_0,\beta_1)}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_0}, \frac{\partial SQ(\beta_0,\beta_1)}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_1} \right),$$

em que

$$\frac{\partial SQ(\beta_0,\beta_1)}{\partial \epsilon_i} = \frac{\partial}{\partial \epsilon_i} \sum_{i=1}^{n} \epsilon_i^2 = 2 \sum_{i=1}^{n} \epsilon_i.$$

$$\frac{\partial \epsilon_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}(y_i - \mu_i) = -1, \quad \frac{\partial \mu_i}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \beta_0 + \beta_1 x_i = 1, \quad \frac{\partial \mu_i}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \beta_0 + \beta_1 x_i = x_i.$$

Vetor gradiente

- Portanto,

$$\begin{aligned}
\nabla SQ(\beta_0,\beta_1) &= \left(-2\sum_{i=1}^{n}\epsilon_i(1); -2\sum_{i=1}^{n}\epsilon_i x_i\right) \\
&= \left(-2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i); -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i\right).
\end{aligned}$$

- Resolver o sistema de equações simultâneas

$$\begin{aligned}
-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i &= 0.
\end{aligned}$$

Vetor gradiente

- Solução

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \bar{y}\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i}.$$

- Graficamente

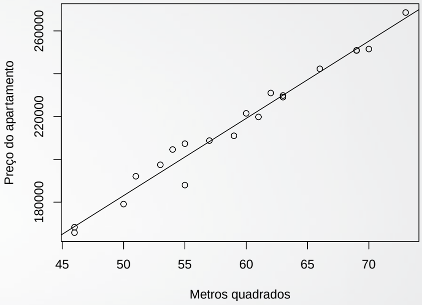


Figura 22. Exemplo regressão linear simples.

Computacionalmente

```r
## Carregando a base de dados
dados <- read.table("data/reglinear.csv",
                    header = TRUE)

## Obtendo beta1
beta1 <- (sum(dados$y*dados$x) -
            mean(dados$y)*sum(dados$x))/
  (sum(dados$x^2) - mean(dados$x)*sum(dados$x))

# Obtendo beta0
beta0 <- mean(dados$y) - beta1*mean(dados$x)
c(beta0, beta1)
```

```
## [1] 2622.752 3608.499
```

```r
## Verificando
coef(lm(y ~ x, data = dados))
```

```
## (Intercept)           x
##    2622.752    3608.499
```

Discussão

Discussão

- Derivadas são essenciais em estatística.
- Maximizar/minimizar funções perda/objetivo.
- O cálculo é por vezes difícil e tedioso.
- Solução de sistemas lineares é tedioso quando possível.
- Álgebra linear ajuda a generalizar as soluções.
- Em situações mais gerais expressões analíticas não serão possíveis de obter.
- Métodos numéricos para resolução de sistemas lineares.
- Métodos numéricos para resolução de sistemas não-lineares.
- Métodos de otimização numérica.

Integrais

Integral indefinida

- Chamamos de **integral indefinida** o oposto ou o inverso da derivada, também chamada de antiderivada.
- A integral indefinida da função $f(x)$ é expressa por

$$\int f(x)dx = F(x) + c.$$

- Exemplo,

$$\int xdx = \frac{x^2}{2} + c,$$

uma vez que se derivarmos $\frac{x^2}{2}$ encontramos x.

Regras de integração

- Integral de uma constante k

$$\int k dx = kx + c.$$

- Integral de 1

$$\int 1 dx = x + c.$$

- Integral da função potência x^n, onde $n \neq -1$

$$\int x^n dx = \frac{1}{n+1} x^{n+1} + c \quad n \neq 1.$$

Regras de integração

- Integral de x^{-1}

$$\int \frac{1}{x} dx = \ln |x| + c \quad \text{se} \quad x \neq 0.$$

- Integral da função exponencial

$$\int a^{kx} dx = \frac{a^{kx}}{k \ln a} + c.$$

- Integral da função exponencial natural

$$\int \exp^{kx} dx = \frac{\exp^{kx}}{k} + c.$$

Regras de integração

- A integral de uma constante multiplicado por uma função é igual à constante multiplicada pela integral da função.

$$\int kf(x)dx = k\int f(x)dx.$$

- A integral da soma é igual a soma das integrais.

$$\int \left[f(x) + g(x)\right] dx = \int f(x)dx + \int g(x)dx.$$

Integral definida

Integral definida

- Suponha que $f(x)$ defina uma curva entre os pontos a e b.
- **Objetivo:** calcular a área sob $f(x)$ entre os pontos $a = 0$ e $b = 1$.

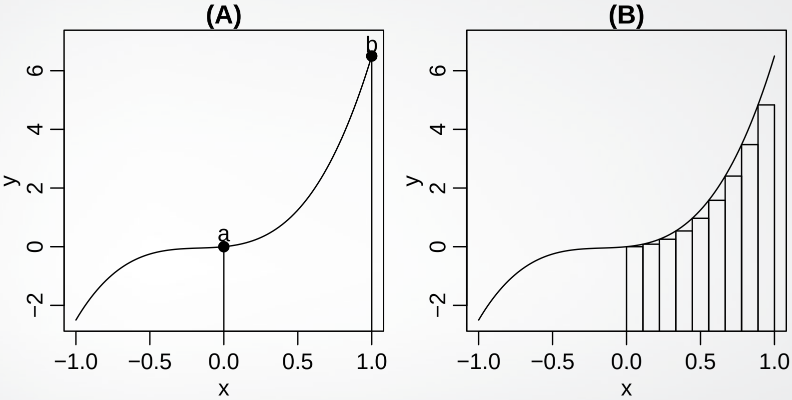


Figura 23. Ilustração de função para o cálculo de área abaixo de uma curva.

Integral definida

- **Ideia:** dividir o intervalo $[a,b]$ em n subintervalos $[a_1,b_1], \ldots, [a_n, b_n]$.
- Em cada subintervalo construir um retângulo.
- Somar as áreas dos retângulos $\sum_{i=1}^{n} f(x_i)\Delta x_i$, onde $\Delta x_i = x_{i+1} - x_i$ é a base do retângulo.
- Aumentar o número de retângulos (ou diminuir sua base).

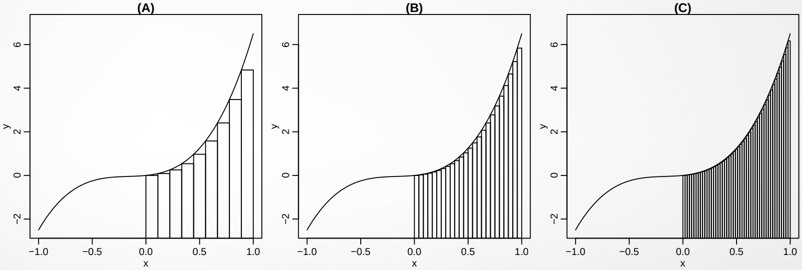


Figura 24. Ilustração da soma de Riemann.

Integral definida

- A área abaixo da curva $f(x)$ pode ser matematicamente expressa por

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\delta x_i = \int_a^b f(x)dx. \tag{4}$$

- **Teorema fundamental do cálculo:** se uma função $f(x)$ é contínua ao longo do intervalo $[a,b]$ e $F(x)$ é a antiderivada de $f(x)$ ao longo desse mesmo intervalo, então

$$\int_a^b f(x)dx = F(b) - F(a).$$

Exemplo

- Cálcule $\int_1^2 x^2 dx$.

$$F(2) - F(1) = \frac{x^3}{3}|_{x=2} - \frac{x^3}{3}|_{x=1} = \frac{8}{3} - \frac{1}{3} = \frac{7}{3} = 2.33\ldots.$$

Soma de Riemann

```r
soma_riemann <- function(n, a, b, fx, ...) {
  intervalos <- seq(a, b, length = n)
  ci <- c()
  soma <- c()
  for(i in 1:c(n-1)) {
    Deltai <- (intervalos[i+1] - intervalos[i]) # Tamanho do intervalo
    ci[i] <- (intervalos[i+1] + intervalos[i])/2 # Ponto central do intervalo
    soma[i] <- fx(ci[i])*Deltai # Cada elemento da soma
  }
  return(sum(soma))
}
soma_riemann <- Vectorize(soma_riemann, "n")
soma_riemann(n = 100, a = 1, b = 2, fx = function(x) x^2)
```

```
## [1] 2.333325
```

Ilustração: soma de Riemann

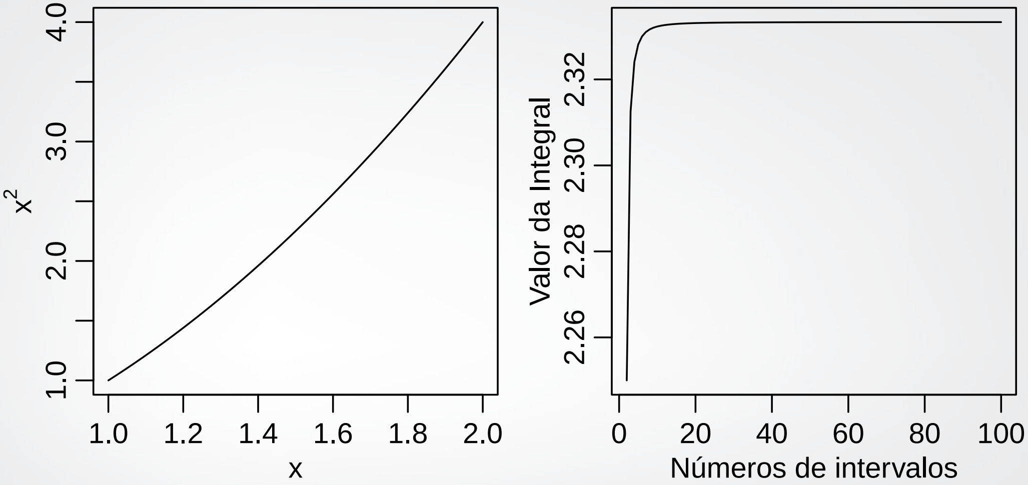


Figura 25. Ilustração da soma de Riemann.

Integração numérica em R

- O R tem uma função nativa para o cálculo de integrais ?integrate.
- Exemplo

```
fx <- function(x) x^2
integrate(fx, lower = 1, upper = 2)
```

```
## 2.333333 with absolute error < 2.6e-14
```

- Outros tipos de integrais
  - Integrais multidimensionais.
  - Integrais impróprias.

Discussão

Discussão

- Integrais são extremamente úteis para obter alguns resultados teóricos em probabilidade.
- Permitem o cálculo de probabilidades para variáveis aleatórias contínuas.
- Técnicas (básicas) de modelagem estatística e *machine learning* não usam integrais diretamente.
- Em geral integrais são mais difíceis de calcular do que derivadas.
- É possível estender a ideia de integrais para funções com duas ou mais variáveis de forma análoga feita para derivadas.
- Integrais em alta dimensão são extremamente difíceis de calcular e/ou aproximar numéricamente.