Introdução ao Cálculo Diferencial Integral para Ciência de Dados

Prof. Wagner Hugo Bonat

Curso de Especialização em
Data Science & Big Data
Universidade Federal do Paraná

7 de março de 2019

Conteúdo

Conteúdo

Funções, limites e continuidade.

Funções

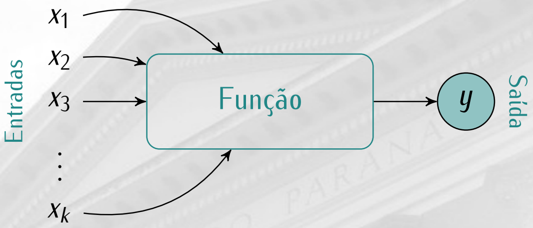
- Definição 1 - Uma **função** escrita como $y = f(x)$ associa um número y a cada valor de x.
- x é chamada de variável **independente**.
- **Domínio de f(x)** é a faixa de valores que x pode assumir.
- y é chamada de variável **dependente**.
- **Imagem de f(x)** é a faixa de valores que y pode assumir.
- Resumindo temos,

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

- O **domínio** e **imagem** de uma função são intervalos.
- Tipos de intervalos:
  - Intervalo aberto **não contêm** as extremidades: Notação (a, b).
  - Intervalo fechado **contêm** as extremidades: Notação $[a, b]$.

Funções: Intuição

Funções: Intuição



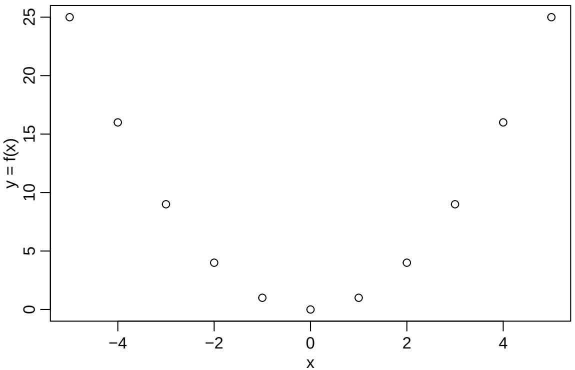
- Exemplos:
  1. Preço do seguro do seu carro;
  2. Concessão de crédito;
  3. Detecção de potenciais clientes;
  4. Detecção de software/operações maliciosas;
  5. etc.

Exemplo

- Considere a função $y = x^2$

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

- Gráfico da função



Funções parametrizadas

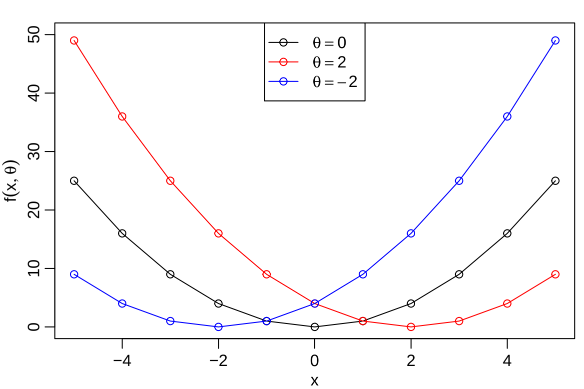
- Definição 2 - **Parâmetro** é uma quantidade conhecida que indexa ou parametriza uma determinada função.
- Em geral os parâmetros mudam o comportamento da função e descrevem quantidades de interesse.
- Notação: $y = f(x; \theta)$, onde θ denota o parâmetro.
- O conjunto de valores que θ pode assumir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

Exemplo: Função parametrizada

- Exemplo: $y = (x - \theta)^2$

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

- Gráfico da função.



Funções com vários parâmetros

- Em geral uma função pode ter vários parâmetros.
- O ideal é que cada parâmetro controle um aspecto da função.
- Exemplo: $y = f(x; \theta_1, \theta_2)$ ou mais geral $y = f(x; \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ é um vetor de parâmetros.
- Função com dois parâmetros:

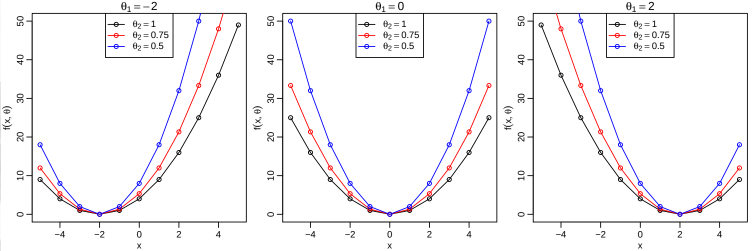
$$y = \frac{(x - \theta_1)^2}{\theta_2}.$$

Funções com múltiplos parâmetros

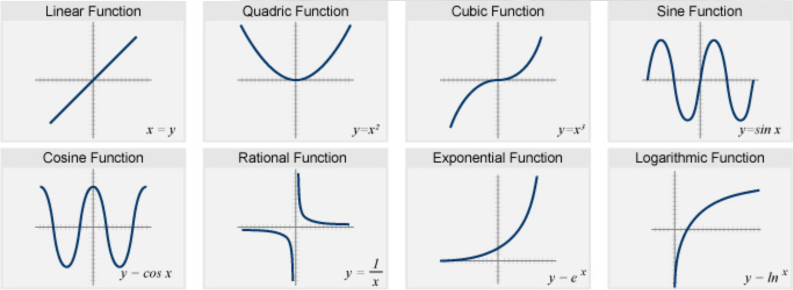
- Função em R.

```
fx = function(x, theta) {
  out <- ((x - theta[1])^2)/theta[2]
  return(out)
}
```

- Gráfico da função para diferentes valores dos parâmetros.



Funções e suas formas

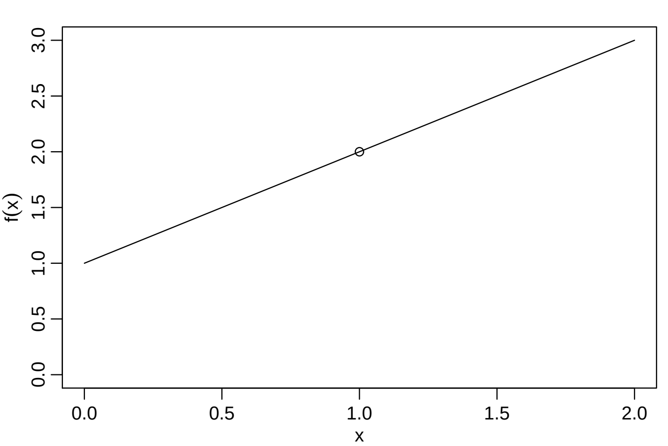


Limite de uma função

- Definição 3 - 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.

Limite de funções - Exemplo 1

- Considere o limite
$$\lim_{x \to 1} (x + 1) = 2.$$



Limite de funções - Exemplo 2

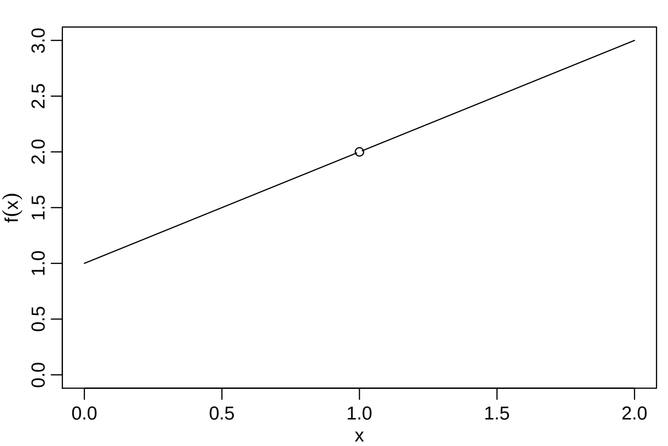
- Considere o limite

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

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

## [1] NaN
```

Limite de funções - Exemplo 2

- Graficamente temos



Limite de funções - Exemplo 2

- 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.

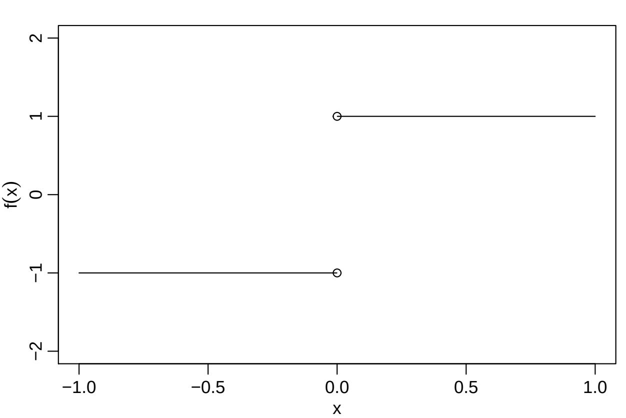
Continuidade de uma função

- Definição 4 - 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 contêm salto vertical.
- Em geral podemos pensar em funções contínuas como sendo funções suaves.

Exemplo - Função não contínua

- Considere a função não continua em 0.

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



Derivadas

Derivada de uma função

- Definição 5 - 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
$$\frac{dy}{dx}\Big|_{x=a} = f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Exemplo: Derivada de uma função

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

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

Interpretação da derivada

- Taxa de mudança instântanea.
- No limite quando $x \to a$ a derivada é a reta tangente ao ponto $(a, f(a))$.
- A reta tangente ao ponto a tem equação dada por $y - f(a) = f'(a)(x - a)$.
- Exemplo: Obtenha a reta tangente a $f(x)$ nos pontos $a = 2$ e $a = -2$.
- Temos $f(x = 2) = -4$ e $f'(x = 2) = -4$, assim

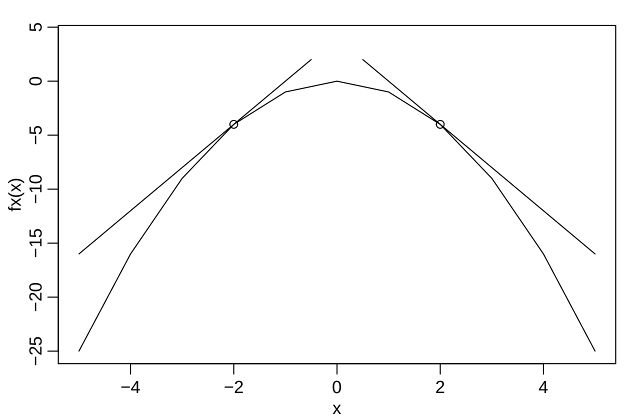
$$\begin{aligned} y - f(x = 2) &= f'(x = 2)(x - 2) \\ y - (-4) &= -4(x - 2) \\ y + 4 = &= -4x + 8 \\ y &= 4 - 4x \end{aligned}$$

Exemplo - Reta tangente a $f(x)$

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

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

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

# Equação da reta y = a + b*x
intercept = (fx(x = 2) - f_prime(x = 2)*2)
intercept

## [1] 4

slope <- f_prime(x = 2)
slope

## [1] -4
```

Exemplo - Reta tangente a $f(x)$



Exemplo - Reta tangente a $f(x)$: Código

```r
par(mar=c(2.6, 2.8, 1.2, 0.5), mgp = c(1.6, 0.6, 0))
x <- seq(-5, 5, l = 11)
plot(fx(x) ~ x, type = "l", ylim = c(-25, 4))
dev_values <- seq(0.5, 5, l = 100)
lines(dev_values, c(intercept + slope*dev_values))
a = (fx(x = -2) - f_prime(x = -2)*-2)
b = f_prime(x = -2)
dev_values <- seq(-5, -0.5, l = 100)
lines(dev_values, c(a + b*dev_values))
points(-2, -4)
points(2, -4)
```

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.
  1. Se $f(x) = \exp(x)$ então $f'(x) = \exp(x)$.
  2. Se $f(x) = \ln(x)$ então $f'(x) = \frac{1}{x}, x > 0$.
- Sejam $f(x)$ e $g(x)$ deriváveis em x e seja c uma constante. Então as funções $f(x) + g(x)$, $cf(x)$ e $f(x) \cdot g(x)$ são deriváveis em x e têm-se
  1. $(f + g)' = f'(x) + g'(x)$.
  2. $(cf)'(x) = cf'(x)$.
  3. $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.

Regras de derivação

- 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()`.
- Exemplos.

Por que derivadas são importantes?

- Derivada é a inclinação (slope) da reta tangente à curva $y = f(x)$.
- **Obtenção de máximo ou mínino de uma função (fundamental !!!).**
- O **máximo** de uma função $f(x)$ é o valor x_n tal que, $f(x_n) \geq f(x), \forall x \in D$.
- O **mínimo** de uma função $f(x)$ é o valor x_1 tal que, $f(x_1) \leq f(x), \forall x \in D$.

Problema: Redução de dados

- Suponha que temos um conjunto de observações y_i para $i = 1, \ldots, n$. Queremos resumir a informação contida em y_i em um único número, digamos μ.
- Problema: Como encontrar μ?

Problema: Redução de dados

- Suponha que temos um conjunto de observações y_i para $i = 1, \ldots, n$. Queremos 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.
- Note que 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**.

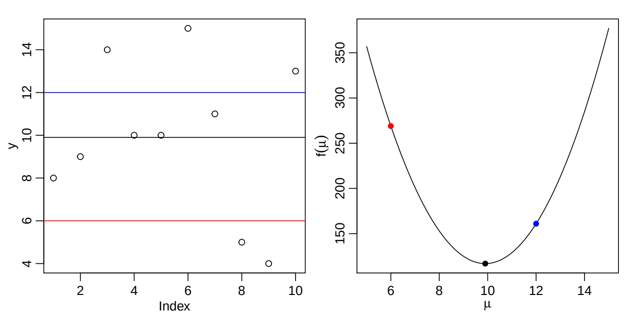
Exemplo: Redução de dados

- Funções em R.

```
set.seed(123)
y <- rpois(10, 10)

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)
}
```

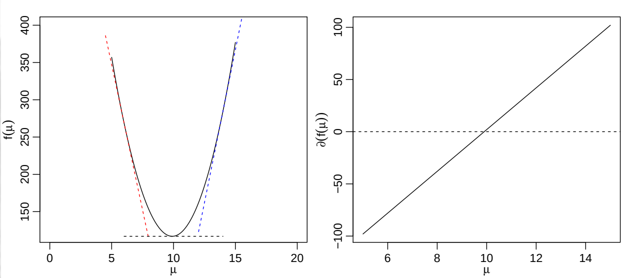
Exemplo: Redução de dados

- Graficamente, temos



Exemplo: Redução de dados

- 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 $f(\mu)$?



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 $f(\hat{\mu}) = 0$.

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

Máximos e mínimos

- Sejam $f(x)$ uma função que admite segunda derivada no intervalo aberto I e $x \in I$.
  1. $f'(x) = 0$ e $f''(x) > 0$ é ponto de mínimo local.
  2. $f'(x) = 0$ e $f''(x) < 0$ é ponto de máximo local.
- $f''(x)$ denota a segunda derivada de $f(x)$, i.e. $\frac{\partial f'(x)}{\partial x}$.
- Em geral em estatística e técnicas padrões de *machine learning* a função objetivo/perda é criada para ter apenas um ponto de mínimo/máximo.
- Em situações patológicas, tais como falta de identificabilidade a função pode ter mais de um mínimo/máximo.

Funções de duas ou mais variáveis indepentes

- Em geral uma função possui apenas uma variável dependente.
- Porém, pode ter duas ou mais variáveis independentes.
- Exemplo 1:

$$z = f(x, y) = \frac{x^2}{2^2} + \frac{y^2}{3^2}.$$

- Exemplo 2: Sejam y_i e x_i quantidades observadas. A equação da reta que relaciona x e y é dada por

$$y_i = f(\beta_0, \beta_1; x_i) = \beta_0 + \beta_1 x_i.$$

Derivadas parciais

- Para uma função $z = f(x, y)$, a derivada parcial de f em relação a x é representada por $\frac{\partial f}{\partial x}$, e é definida por

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h},$$

desde que o limite exista.

- De forma similar, a derivada parcial de $f(x, y)$ em relação a y é representada por

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}.$$

- De forma geral, $\frac{\partial f(x)}{\partial x}$ é obtida derivando $f(x)$ considerando y fixo e vice-versa.

Gradiente e Hessiano

- O gradiente de uma função $f(x, y)$ é o vetor composto pelas derivadas primeira de $f(x, y)$ em relação a x e y, i.e.

$$\nabla f(x, y) = \left(\frac{\partial f(x, y)}{\partial x}, \frac{\partial f(x, y)}{\partial y} \right).$$

- O Hessiano de uma função $f(x, y)$ é a matrix

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial f(x,y)}{\partial x \partial y} \\ \frac{\partial f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial x^2} \end{pmatrix}.$$

- As definições se estendem naturalmente para mais de duas variáveis.
- Vamos revisar as ideias de vetores e matrizes (álgebra linear).

Expansão de funções em Série de Taylor

- Aproximação por Série de Taylor é fundamental em estatística e métodos numéricos.
- É uma forma simples de obter o valor aproximado de uma função perto de um ponto conhecido x_0.
- Dada uma função $f(x)$ derivável $(n+1)$ vezes em um intervalo contendo um ponto $x = x_0$, temos

$$
\begin{aligned}
f(x) &= f(x_0) + (x - x_0)\frac{df(x)}{dx}\Big|_{x=x_0} + \frac{(x-x_0)^2}{2!}\frac{d^2 f(x)}{dx^2}\Big|_{x=x_0} + \\
\dots &+ \frac{(x-x_0)^n}{n!}\frac{d^n f(x)}{dx^n}\Big|_{x=x_0} + R_n(x).
\end{aligned}
$$

- Aproximação similar é possível para funções com duas ou mais variáveis.

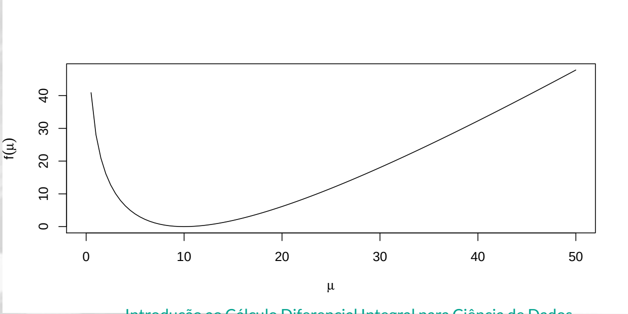
Expansão em série de Taylor

- Considere a seguinte função

$$f(\mu) = 2\left(10 \log \frac{10}{\mu} - 10 + \mu\right).$$

Aproxime ao redor do ponto $\mu_0 = 10$.

```
f <- function(mu) {
  out <- 2*(10*log(10/mu) - 10 + mu)
  return(out)
}

plot(f, 0, 50, ylab = expression(f(mu)), xlab = expression(mu))
```



Expansão em série de Taylor

- Primeira derivada

$$f'(\mu) = 2\left(1 - \frac{10}{\mu}\right).$$

- Segunda derivada

$$f''(\mu) = \frac{20}{\mu^2}.$$

- Aproximação em série de Taylor (segunda ordem) ao redor de μ_0.

$$f(\mu) \approx f(\mu_0) + (\mu - \mu_0)f'(\mu = \mu_0) + \frac{(\mu - \mu_0)^2}{2!}f''(\mu = \mu_0).$$

Expansão em série de Taylor

- Aproximação de Taylor (genérica)

```
taylor_ap <- function(mu, mu0, f, f_prime, f_dprime) {
  app <- f(mu = mu0) + (mu - mu0)*f_prime(mu = mu0) +
    (((mu - mu0)^2)/(2))*f_dprime(mu = mu0)
  return(app)
}
```

- Derivadas

```
f_prime <- function(mu) {2*(1- 10/mu)}
f_dprime <- function(mu) {20/mu^2}
```

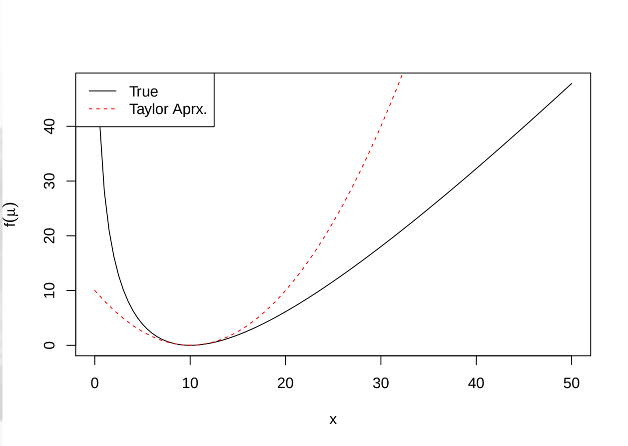
- Aproximando a função

```
taylor_ap(mu = c(9,10,11), mu0 = 10, f = f,
          f_prime = f_prime, f_dprime = f_dprime)

## [1] 0.1 0.0 0.1
```

Expansão em série de Taylor

- Graficamente, temos

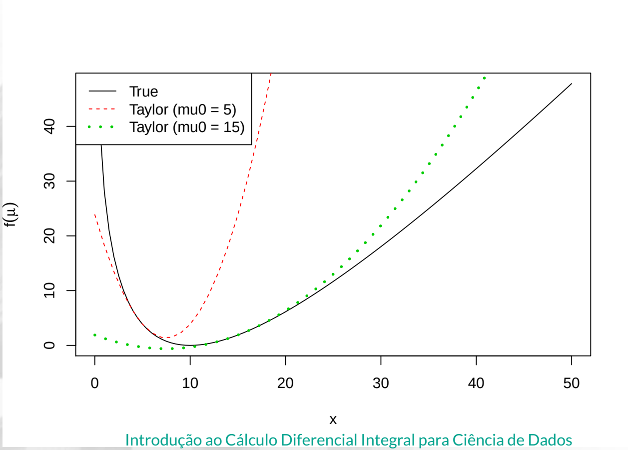
```
plot(f, 0, 50, ylab = expression(f(mu)), xlab = "x")
lines(0:50, taylor_ap(mu = 0:50, mu0 = 10, f = f, f_prime = f_prime,
                      f_dprime = f_dprime), col = "red", lty = 2)
legend("topleft", legend = c("True","Taylor Aprx."), lty = c(1,2), col = c(1,2))
```



Expansão em série de Taylor

- Graficamente, temos

```
plot(f, 0, 50, ylab = expression(f(mu)), xlab = "x")
lines(0:50, taylor_ap(mu = 0:50, mu0 = 5, f = f, f_prime = f_prime,
                      f_dprime = f_dprime), col = "red", lty = 2)
lines(0:50, taylor_ap(mu = 0:50, mu0 = 15, f = f, f_prime = f_prime,
                      f_dprime = f_dprime), col = 3, lty = 3, lwd = 3)
legend("topleft", legend = c("True","Taylor (mu0 = 5)","Taylor (mu0 = 15)"),
       lty = c(1,2,3), col = c(1,2,3), lwd = c(1,1,3))
```



Exemplo - Regressão linear simples

- Seja $y_i (i = 1, \ldots, n)$ observações de alguma variável de interesse.
- Seja x_i uma outra variável que queremos relacionar com y_i através de um reta, i.e.

$$y_i = \beta_0 + \beta_1 x_i.$$

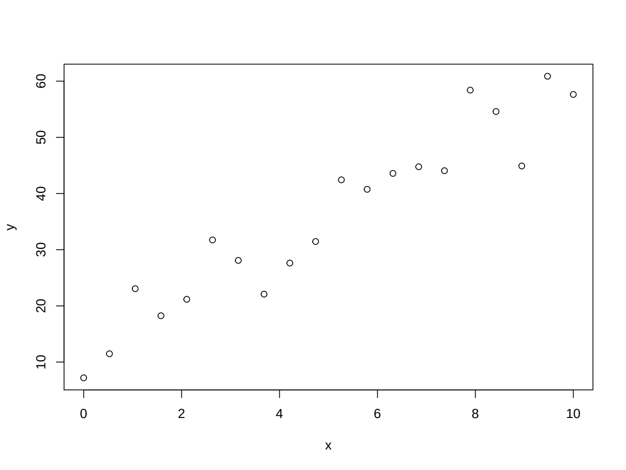
- Problema: Encontrar β_0 e β_1 tal que

$$SQ = f(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2,$$

seja a menor possível.

Exemplo - Regressão linear simples

- Graficamente, temos



Como encontrar β_0 e β_1 que minimizam a SQ?

- Abordagem 1
  1. Obter o vetor gradiente

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

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

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

- Abordagem 2 - Obter a derivada analiticamente, mas resolver o sistema numericamente.
- Abordagem 3 - Obter a derivada e resolver o sistema numericamente.
- Abordagem 4 - Usar um algoritmo de otimização genérico.

Vetor gradiente

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

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

onde

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

$$\frac{\partial \epsilon_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} (y_i - \mu_i) = -1.$$

$$\frac{\partial \mu_i}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \beta_0 + \beta_1 x_i = 1.$$

$$\frac{\partial \mu_i}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \beta_0 + \beta_1 x_i = x_i$$

Vetor gradiente

- Vetor gradiente,

$$\nabla f(\beta_0, \beta_1) = \left(-2 \sum_{i=1}^{n} \epsilon_i (1); -2 \sum_{i=1}^{n} \epsilon_i x_i \right)$$

$$= \left(-2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i); -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i \right).$$

- Resolver o sistema de equações:

$$-2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (1)$$

$$-2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (2)$$

Resolvendo o sistema de equações

- Pela Eq. (1) temos,
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{3}$$
- Substituindo Eq.(3) em Eq. (2) e resolvendo em $\hat{\beta}_1$, temos

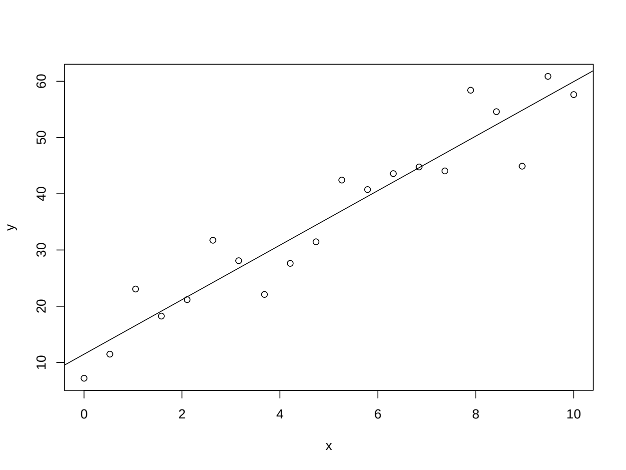
$$\hat{\beta}_1 = \frac{\bar{y} \sum_{i=1}^{n} - \sum_{i=1}^{n} y_i x_i}{\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}.$$

Implementação em R

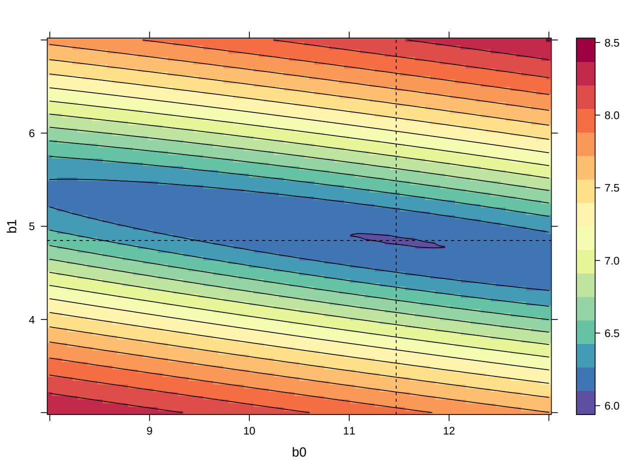
- Simulando um conjunto de dados

```
set.seed(123)
b0 = 10
b1 = 5
x <- seq(0,10, l = 20)
mu <- b0 + b1*x
y <- rnorm(20, mean = mu, sd = 5)
head(cbind(y, x))

##              y         x
## [1,]  7.197622 0.0000000
## [2,] 11.480691 0.5263158
## [3,] 23.056699 1.0526316
## [4,] 18.247279 1.5789474
## [5,] 21.172754 2.1052632
## [6,] 31.733220 2.6315789

b1 <- (mean(y)*sum(x) - sum(y*x))/(mean(x)*sum(x) - sum(x^2))
b0 <- mean(y) - b1*mean(x)
c(b0,b1)

## [1] 11.470239  4.847576

# Conferindo
coef(lm(y ~ x))

## (Intercept)           x
##   11.470239    4.847576
```

Visualizando a reta ajustada



Visualizando a superfície objetivo



Discussão

- Derivadas são essenciais em *data science*.
- 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 sistema não-lineares.
- Métodos numéricos de otimização.

Integrais

Integral de uma função

- Integral indefinida: É o oposto ou inverso da derivada, também chamada de **antiderivada**.
- Sendo $g(x)$ a derivada de $f(x)$, então a antiderivada ou integral indefinida de $g(x)$ é $f(x)$, sendo escrita como:

$$f(x) = \int g(x)dx.$$

- Exemplo: $\int xdx = \frac{x^2}{2}$. Uma vez que,

$$\frac{\partial}{\partial x}\frac{x^2}{2} = \frac{2x^{2-1}}{2} = x.$$

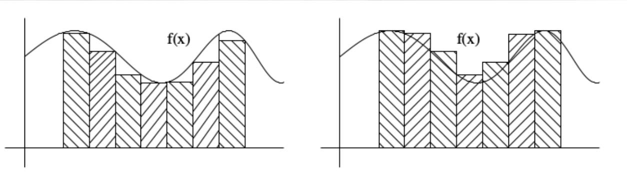
Integral de uma função

- Integral definida I é denotado por
$$I = \int_a^b f(x)dx.$$
- A integral é um número, e existe em um intervalo fechado $[a, b]$.
- Considere uma função $f(x)$ definida e contínua no intervalo $[a, b]$.
- O domínio de $f(x)$ pode ser dividido em n subintervalos definidos por $\Delta x_i = x_{i+1} - x_i$, onde $i = 1, \ldots, n$.
- Soma de Riemann
$$\sum_{i=1}^{n} f(c_i)\Delta x_i$$
onde c_i é um número pertecente ao subintervalo $[x_i, x_{i+1}]$.

Integral definida

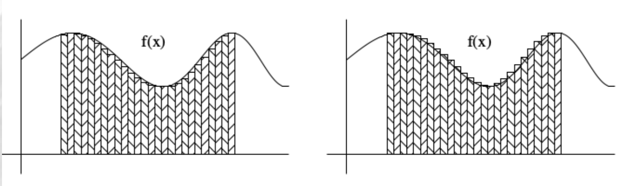
- A integral definida é obtida como o limite da soma de Riemann quando a largura de todos os subintervalos de $[a, b]$ tende a zero:

$$I = \int_a^b f(x)dx = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i.$$



Integral definida

- Aumentando o número de retângulos a área é melhor aproximada.
- Quando $n \to$ ou $\Delta x_i \to 0$ a soma de Riemann corresponde a área abaixo da curva $f(x)$.



- Podemos facilmente imitar essa ideia numericamente em R.

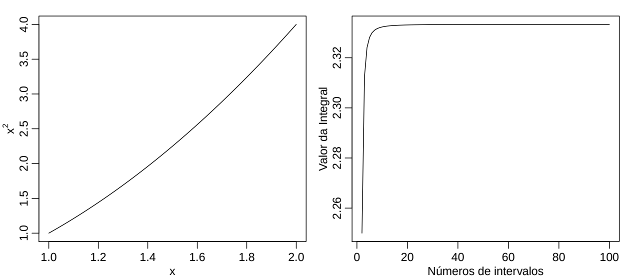
Soma de Riemann

- Função para calcular a soma de Riemann.

```
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 de Riemann

- Calculando a área abaixo da curva $f(x) = x^2$ em $[1, 2]$.



- Valor aproximado com $n = 100$.

```
soma_riemann(n = 100, a = 1, b = 2, fx = fx)

## [1] 2.333325
```

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
$$\int_a^b f(x)dx = F(b) - F(a).$$
- Existem tabelas de antiderivadas.
- Por exemplo, sendo n um natural ($n \neq 1$) temos
$$\int x^n dx = \frac{x^{(n+1)}}{n+1}.$$
- Assim, temos
$$\int_1^2 x^2 dx = F(2) - F(1) = \frac{x^3}{3}|_{x=2} - \frac{x^3}{3}|_{x=1} = \frac{8}{3} - \frac{1}{3} = \frac{7}{3} = 2.33\ldots$$

Propriedades e Regras básicas de Integração

- Sejam $f(x)$ e $g(x)$ funções contínua em um certo intervalo e k uma constante. São válidas as propriedades:
  1. $\int kf(x)dx = k \int f(x)dx$.
  2. $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$.
  3. $\int f'(x)g(x)dx = f(x)g(x) - \int f(x)g'(x)$.
- Regras básicas de integração:
  1. $\int kdx = kx + c$.
  2. $\int x^n dx = \frac{x^{(n+1)}}{n+1}$.
  3. $\int \frac{1}{x}dx = \ln|x| + c$.
  4. $\int \exp x dx = \exp x + c$.
  5. $\int a^x dx = \frac{a^x}{\ln(a)} + c$.
- Ver tabela no material suplementar.
- Lista de exercícios.

Discussão

- Integrais são extremamente úteis para obter alguns resultados teóricos em probabilidade.
- 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 numericamente.