04_2-Caderno-InfEst-parte2

Helena R. S. D'Espindula

2024-04-19

Cálculo Diferencial e Integral

Problemas convencionais em ciência de dados

- Problemas convencionais em ciência de dados
- Previsão ou predição → O que vai acontecer?
- Classificação → Qual o tipo de um determinado objeto?
- Agrupamento → Qual a melhor forma de agregar objetos?
- Prescrição → O que devo fazer?
- 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

- 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,

$$\frac{x \in Dominio}{Independente} \longrightarrow f(x) \longrightarrow \frac{x \in Imagem}{Dependente}$$

- 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: $ {N} = {0, 1, 2, 3...} $ {N} = {0,1,2,3, ... }$.
    - Inteiros: $Z = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$
    - Racionais $Q = ab|a, b \in Z, b! = 0$
    - Irracionais: Conjunto de números que não são racionais.
    - Reais: União de todos os números mencionados acima, notação {R}.

- Distinção importante R (double) e Z (integer).

- Considere a função $y = x^2$.

- Em R temos

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

- Avaliando a função em alguns pontos.

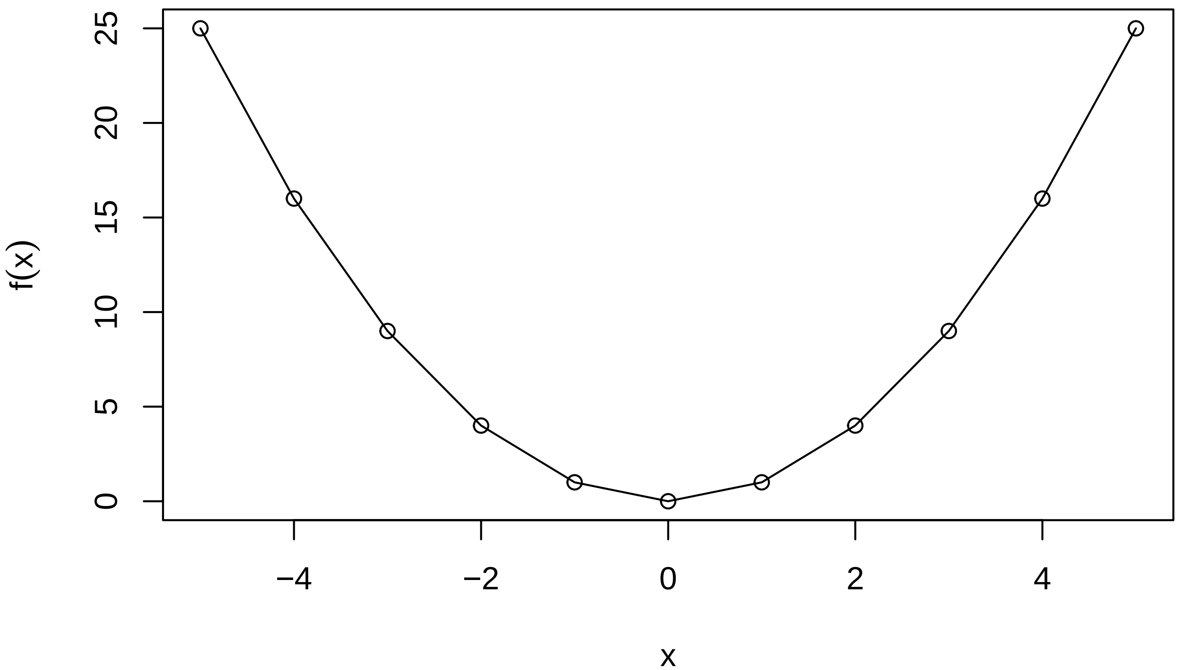
```
x_vec <- c(-5, -4, -3, -2, -1,
0, 1, 2, 3, 4, 5) #concatenação
minha_funcao(x = x_vec) #automaticamente vetorizado
```

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

Funções unidimensionais

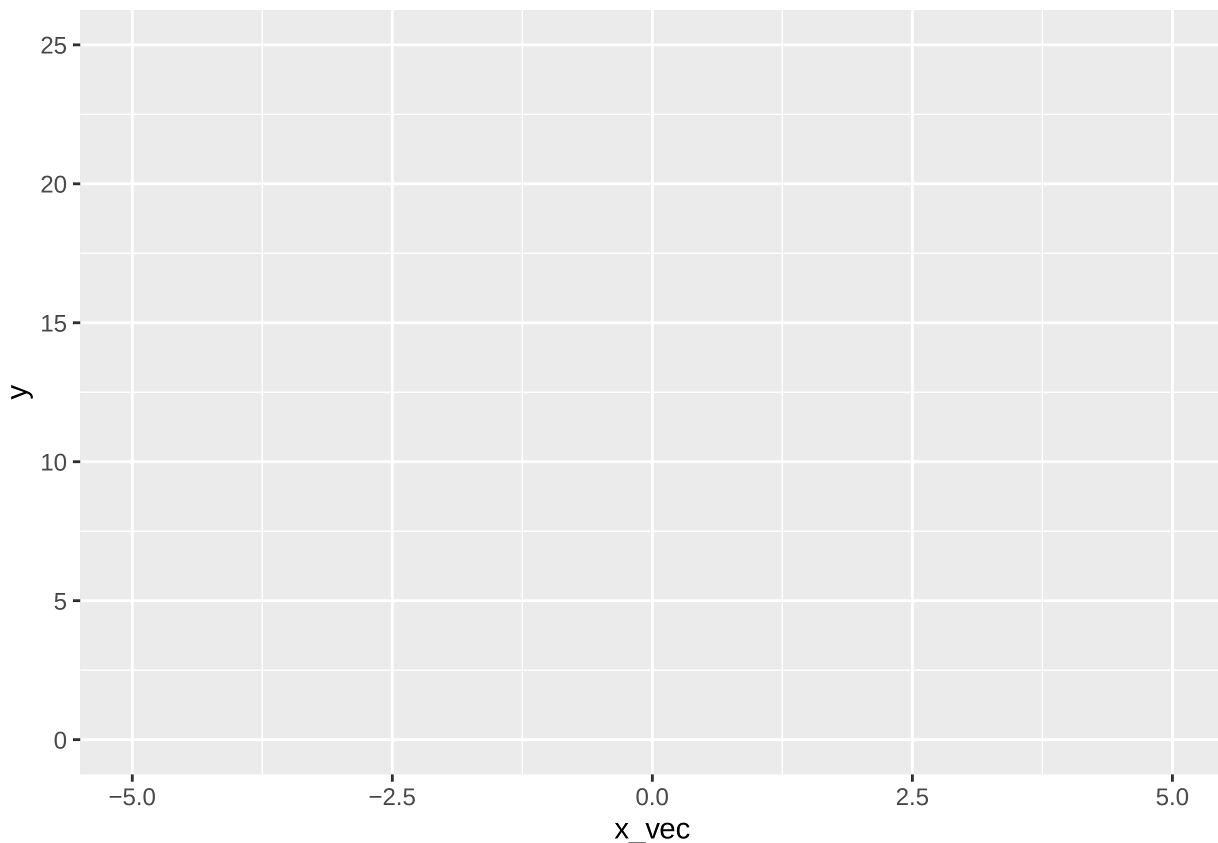
- Uma função $y = f(x)$ é dita ser de apenas uma variável (unidimensional). Ou seja, só uma entrada
- Pode ser desenhada em um espaço bidimensional, o chamado R_2. Gafico de x e y
- O espaço R_2 é formado por todas as duplas ordenadas de valores reais.
- A variável dependente y é representada no eixo vertical.
- A variável dependente x é representada no eixo horizontal.

```
## 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)
```




```r
## ou com GGPLOT

ggplot(mapping = aes(x_vec,y))
```



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 (theta minusculo) pode assumir é chamado de espaço paramétrico (theta maiusculo).
- Notação

$$\theta \in \Theta$$

- Exemplo: $y = (x - \theta)^2$. Theta joga o grafico mais para direita ou esquerda
- Computacionalmente:

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

```
## Criar grafico
```

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

```
## Criar grafico
```

Declividade

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

```
## Criar grafico
```

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. (**Atenção**: x é um vetor nesse caso!) - $x = (x_1, \dots, x_p)^T$ 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 $x = (0, 0)^T$, $x = (3, 0)^T$ e desenhe seu gráfico. - Avaliando nos pontos

[[**ARRUMAR**]] $y = \sqrt{25 - 02 - 02} = 5$ e $y = \sqrt{25 - 32 - 02} = 4$

Computacionalmente - Implementação computacional

```
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
```

```
## [1] 5
fx1x2(x = entrada2)
```

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

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

- Avaliando uma função bidimensional.

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

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

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

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

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

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

- O gráfico da função é o conjunto das triplas ordenadas (y, x1, x2) que satisfazem a função.

Passo-a-passo para desenhar funções bidimensionais

- Neste caso estamos no espaço R_3.
- (A) Montar uma grade de valores combinando valores para x1 com valores para x2.
- (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. (poderia ser uma topografia, seria uma hemi-esfera)

```
## Fazer grafico
```

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

Gráficos bidimensionais - Em geral usamos uma grade mais precisa.

```
# Fazer grafico
```

Figura 6. Ilustração do gráfico de uma função de duas variáveis de entrada.(curva de nivel ou iso-linha)

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.

Funções polinômiais

- Funções polinômiais são funções do tipo

$$y = \beta_0 + \beta_1 x + \beta_2$$

$$x^2 + \ldots \beta_p x^p$$

- 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:

```
# Fazer grafico
```

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. 1.

$$xa(xc) = xa + c;$$

2.

$$(xa)c = xac;$$

3.

$$3.(xz)a = xa(za);$$

4.

$$4.(xz)c = xczc;$$

5.

$$5.1xa = x - a;$$

6.

$$6.xaxc = xa - c;$$

7.
$$7.\sqrt{x} = \frac{x^1}{2};$$
8.
$$8.a\sqrt{x} = x1/a;$$
9.
$$9.c\sqrt{xa} = \frac{xa}{c}.$$

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

$$\lim_{n\to\infty}(1 + \frac{1}{n})n = 2.718281828$$

- Propriedades importantes:

1.
$$e0 = 1.$$
2.
$$e1 = e = 2.71828$$
3.
$$e(eb) = ea + b.$$
4.
$$(ea)b = eab$$
5.
$$eaeb = ea - b.$$

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 log10.
- 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, a potência à qual a deve ser elevado, para obter-se y.

[[**ARRUMAR**]] log a $y = x x é

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 $\ln(y) = t + 1$ **Observação**: $e = 2.718281828459045 =$ número de Euler.

Outras funções de interesse

- Sigmóide ou logística: $y = 11 + e - x$
- Tangente hiperbólica: $y = ex - e - xex + e - x$.
- Linear retificada (ReLU): $y = max0, x$. (maximo entre 0 e x?. Vale 0 até o 0 e depois "sobe")
- Leaky ReLU: $y = max\alpha x, x$, onde α é uma parâmetro conhecido.

##Desenho do gráfico das funções

```
#desenho aqui
```

Normal

$$y = (x - 0)^2/\theta$$

$$exp{-}(n - \theta_1)^2/\theta_2$$

```
## fazer grafico
```

[[**ARRUMAR**]] .....normal

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

$$\lim_{x \to 1} x^2$$

[[**ARRUMAR**]]

- Computacionalmente

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

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

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

```
## Figura 11. Desenho do gráfico da função
```

Exemplo - Note que

[[**ARRUMAR**]]

- Definição intuitiva: o limite de uma função é o valor que achamos natural para ela em um determinado ponto.
- Essa função não é continua (no ponto)

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.(mudanças suaves, ou não abruptas)
- 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.

$$\lim_{x \to 0} \frac{|x|}{x} = \{-1 x < 0 e 1 x > 0$$

```
## 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

$$\lim_{x \to p} [f(x) + g(x)] = L_1 + L_2$$

$$\lim_{x \to p} k f(x) = k$$

$$\lim_{x \to p} f(x) = kL_1$$

$$\lim_{x \to p} f(x)g(x) = \lim_{x \to p} f(x) * \lim_{x \to p} g(x) = L_1 * L_2$$

$$\lim_{x \to p} f(x)g(x) = L_1 * L_2$$

, desde que $L_2 \neq 0$.

Derivadas

- 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

$$\frac{dy}{dx}$$

ou

$$y'$$

ou

$$\frac{df}{dx}$$

ou

$$f'(a)$$

é o valor

$$\frac{dy}{dx}|x = a = f'(a)$$

[[**ARRUMAR**]]

$$\lim_{h \to 0} = \frac{f(a+h) - f(a)}{h}$$

- Interpretação da derivada
- Taxa de mudança instântanea.
- No limite quando $x \longrightarrow a$ a derivada é a reta tangente ao ponto $(a, f(a))$.
- Equação da **reta tangente** ao ponto a: $y - f(a) = f'(a)(x - a)$.(coeficiente angular é $\beta_1 - > y = \beta_0 + \beta_1 * x$) [[**ARRUMAR**]]

Exemplo

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

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

$$= \lim_{h \to 0} \frac{-(x+h)^2 - (-x2)}{h}$$

$$= \lim_{h \to 0} \frac{-(x^2 + 2xh + h^2) + x^2}{h}$$

$$= \lim_{h \to 0} \frac{-x^2 - 2xh - h^2 + x^2}{h}$$

$$= \lim_{h \to 0} \frac{-2xh - h^2}{h}$$

$$= \lim_{h \to 0} -2x - h = -2x$$
$$= f'(x) = -2x$$

[[CONFERIR]]

Obtenha a reta tangente a $f(x)$ nos pontos $x = 2$ e $x = -2$. - Temos:

$$f(x = 2) = -4$$

$$f'(x = 2) = -4$$

assim

$$y - f(x = 2) = f'(x = 2)(x - 2)$$
$$y - (-4) = -4(x - 2)$$
$$y + 4 == -4x + 8$$
$$y = 4 - 4x$$

- Computacionalmente - 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)
slope <- f_prime(x = 2)
c(intercept, slope)
```

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

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

```
## Figura 14. Desenho de uma função e retas tangentes.
```

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 $f(x) = xn$ então $f'(x) = n * n^-1$
3. Se $f(x) = x - n$ então $f'(x) = -n * -n^{-1}$

4. Se $f(x) = \frac{x_1}{n}$ então $f'(x) = \frac{1}{n} * \frac{x_1}{n} - 1$

- Derivada de funções especiais

5. Se $f(x) = exp(x)$ então $f'(x) = exp(x)$.
6. Se $f(x) = ln(x)$ então $f'(x) = 1x, x > 0$

- 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. $(f * g)'(x) = f'(x)g(x) + f(x)g'(x)$.
10. $(fg)'(x) = f'(x)g(x) - f(x)g'(x)[g(x)]2$.

- Exemplo: obtenha a derivada de $f(x) = 2 + 3x$
- Solução: $f'(x) = 3$
- Computacionalmente

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

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

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

Regra da cadeia

- "Uma função dentro da outra"
- Sejam $y = f(x)$ e $x = g (t) $ duas funções deriváveis, com $I \in Df$. A função composta $h(t) = f(g(t))$ é derivável, sendo $h'(t) = f'(g(t))g'(t), t \in Dg$.
- 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 (derivada de sen e cos)
$$in(g(x))$$
, onde
$$g(x) = 2x^3 - 4x$$
2. Usando a regra da cadeia temos:
$$f'(g(x)) = cos(2x^3 - 4x)$$
and
$$g'(x) = 6x^2 - 4.$$
3. Assim, a derivada fica dada por
$$cos(2x^3 - 4x) * (6x^2 - 4).$$
4. Computacionalmente

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

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

```
## cos(2 * x^3 - 4 * x) * (2 * (3 * x^2) - 4)
## ou cos(2x^3 - 4x) * (6x^2) - 4)

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

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

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

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: $fracdnydxn$ que é interpretada como a n-ésima derivada de y em relação a x
- Exemplo: obtenha as derivadas até a ordem 5 da função $y = 2x^4 + 5x^3 + 2x^2$ $dy = 2x^4 + 5x^3 + 2x^2$ $dx = 8x^3 + 15x^2 + 4x$ d2y $dx2 = 24x^2 + 30x + 4$ d3y $dx3 = 48x + 30$ d4y $dx4 = 48$ e d5y $dx5 = 0$ daqui em diante é zero

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)$4 esteja definida, tal que $f(c) >= f(x)$ 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. (maximo e minimo dentro de um trecho de grafico)

```
## Figura 15. Ilustração de máximo/mínimo relativos.
```

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

Pontos extremos (pico do morro ou fundo do vale)

- 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).

```
## Fazer figura
```

Figura 17. Ilustração de uma função com a reta tangente.ponto extremo tem inclinação zero (a3 na figura)

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?

- 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 $f(\mu) = \sum_{i=1}^{n}(y_i - \mu)^2$ (soma de quadrados da diferença de cada valor para media, ou seja o quanto eu perdi ao trocar os y por μ), seja a menor possível.
- Uma vez que temos os números observados y_i a única quantidade desconhecida éμ.
- 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.
- dervar e igualar a 0.

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, 0, 8), y = y)
```

```
## [1]  117  161 1097  153
```

```
## [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

$$f(y) = \sum_{i=1}^{n}(y_i - \mu)^2$$

- 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 $f'(\mu\hat{)} = 0$
- Assim, temos (regra da cadeia!!)

$$f(y) = \sum_{i=1}^{n}(y_i - \mu)^2$$
$$\varepsilon_i = y_i - \mu$$
$$f'(\mu) = \sum_{i=1}^{n}(\varepsilon_i)^2$$
$$f'(\mu) = 2\sum_{i=1}^{n}(y_i - \mu)\frac{d}{d\mu}(y_i - \mu)$$
$$f'(\mu) = 2\sum_{i=1}^{n}(y_i - \mu)(-1)$$
$$f'(\mu) = -2\sum_{i=1}^{n}(y_i - \mu)$$

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

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

$$-2\sum_{i=1}^{n}(y_i - \hat{\mu}) = 0$$
$$-\sum_{i=1}^{n}(y_i - n\hat{\mu}) = 0$$
$$n\hat{\mu} = \sum_{i=1}^{n} y_i$$
$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}$$

OU SEJA MÉDIA!!!

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

- 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 $x1$, quando a outra variável independente $x2$ é mantida constante.
- Como obter a derivada parcial?
- A derivada parcial em relação a $x1$ é obtida derivando $f(x1, x2)$ "fingindo" que $x2$ é uma constante.
- A derivada parcial de $f(x1, x2)$ em relação a $x2$ é obtida derivando $f(x1, x2)$ mantendo $x1$ 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 $y = 5x_1^3 + 3x_1x_2 + 4x_2^2$

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

ou

```
D(expression( 5 * x^3 + 3 * x * c + 4* c^2 ), name = "x")
```

```
## 5 * (3 * x^2) + 3 * c
```

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

```
D(expression( 5 * x^3 + 3 * x * c + 4* c^2 ), name = "c")
```

```
## 3 * x + 4 * (2 * c)
```

Derivadas parciais de ordem superior - Derivadas parciais de segunda ordem

$$\partial 2f(x1, x2)$$

$$\partial x21$$

e

$$\partial 2f(x1, x2)$$

$$\partial x22$$

indica que a função foi diferenciada parcialmente em relação a $x1$ ou $x2$ duas vezes. - Derivada parcial cruzada (ou mista)

$$\partial 2f(x1, x2)$$

$$\partial x1\partial x2$$

indica que primeiro derivamos em $x1$ e depois em $x2$. - A ordem da derivada cruzada não importa (se ambas contínuas), ou seja

$$\partial 2f(x1, x2)$$

$$\partial x1\partial x2$$

$$\partial 2f(x1, x2)$$

$$\partial x2\partial x1$$

Exemplo: derivadas parciais de segunda ordem - Obtenha as derivadas parciais de até segunda ordem em relação a x_1 e x_2 de $y = 7x31 + 9x1x2 + 2x52$

- Derivadas parciais de primeira ordem

$$\frac{\partial y}{\partial x_1} = 21x21 + 9x2, \partial y \partial x2 = 9x1 + 10x42$$

- Derivadas parciais de segunda ordem (segunda derivadas direta) [[**ARRUMAR**]]

- Derivadas parciais de segunda ordem (termos cruzados) [[**ARRUMAR**]]

- As cruzadas dão sempre iguais.

Máximos e mínimos funções muldimensionais

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

Exemplo

Considere a função $y = 6x21$ $-9x1 - 3x1x2 - 7x2 + 5x22$

Encontre os pontos críticos e determine se são de máximo ou minímo. - Graficamente

- Calcular as derivadas parciais de primeira ordem da função [[**ARRUMAR**]]

- Derivando em x 1, temos [[**ARRUMAR**]]

- Derivando em x 2, temos [[**ARRUMAR**]]

- Resolver o sistema de equações $12x1 - 9 - 3x2 = 0$ $-3x1 - 7 + 10x2 = 0$

- Solução: $x_1 = 1$ e $x_2 = 1$. ## USA NA PROXIMA

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

[[**ARRUMAR**]]

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

[[**ARRUMAR**]]

Assim, temos que [[**ARRUMAR**]]

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

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 [[**ARRUMAR**]]
- E para o caso multidimensional [[**ARRUMAR**]]

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: [[**ARRUMAR**]]

onde o termo R n(x) é chamado de resíduo ou erro, e dado por [[**ARRUMAR**]]

sendo E 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 [[**ARRUMAR**]]
- Aproximação de segunda ordem [[**ARRUMAR**]]

Graficos

- Quanto mais se afasta de zero pior fica a aproximação.
- A Tailor em n+1 graus é igual a original

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.

Grafico

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, ... ,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 ~ tamanho

$$Preço = f(metroquadrado)$$

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

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

$$f^*(x) = f^*(x_0) + (x - x_0)f^{*'}(x) + R_n(x)$$

- Ignorando o termo residual R n(x) [[**ARRUMAR**]]
- Rearranjando os termos obtemos [[**ARRUMAR**]]
- De forma equivalente, temos
$$yi = \beta_0 + \beta_1 xi + Rn(xi),$$
em que o termo R n(x i) é o erro cometido em aproximar y i por $\beta_0 + \beta_1 xi$
- Notação usual $Ei = yi - (\beta_0 + \beta_1 xi)$.
- 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 [[**ARRUMAR**]]
- Obter o vetor gradiente [[**ARRUMAR**]]
- Encontrar $\hat{\beta_0}$ e $\hat{\beta_1}$ tal que [[**ARRUMAR**]]

1. Chame $y_i - (\beta_0 + \beta_1 xi) = erro_i$.
2. Chame $\beta_0 + \beta_1 xi = \mu i$
3. Assim, [[**ARRUMAR**]]

Vetor gradiente - Portanto, [[**ARRUMAR**]]

- Resolver o sistema de equações simultâneas (derivadas de beta 0 e beta 1) [[**ARRUMAR**]]
- Solução [[**ARRUMAR**]]

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

```
##           y  x
## 1  207317.7 55
## 2  250845.7 69
## 3  165755.0 46
## 4  219816.7 61
## 5  268582.2 73
## 6  229059.7 63
## 7  179098.1 50
## 8  168301.8 46
## 9  250917.7 69
## 10 230990.3 62
## 11 229818.7 63
## 12 221484.1 60
## 13 187977.0 55
## 14 251526.9 70
## 15 242289.5 66
## 16 211007.7 59
## 17 197455.5 53
## 18 192100.8 51
## 19 208762.3 57
## 20 204593.1 54
```

```r
## 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
## [1] 2622.752 3608.499
## Verificando
```

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

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

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

Modelo: $\hat{y} = 2622.752 + 3608.499*$ metrosquadrados

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.

Aula 2024-04-06

Integrais

Anti-derivada

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,

$$\int xdx = x2$$

  2 + c , uma vez que se derivarmos x 2 2 encontramos x

[[**COMPLETAR**]]

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 = 2, a = 1, b = 2, fx = function(x) x^2)
```

```
## [1] 2.25
```

```r
soma_riemann(n = 10, a = 1, b = 2, fx = function(x) x^2)
```

```
## [1] 2.332305
```

```r
soma_riemann(n = 50, a = 1, b = 2, fx = function(x) x^2)
```

```
## [1] 2.333299
```

```r
soma_riemann(n = 100, a = 1, b = 2, fx = function(x) x^2)
```

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

```r
soma_riemann(n = 1000, a = 1, b = 2, fx = function(x) x^2)
```

```
## [1] 2.333333
```

Integração numérica em R

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

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

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

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

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

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.