

Lista 1 — Revisão de Pré-requisitos

Bases 5 — Ciência de Dados · UnDF · 2026.2

Prof. Douglas Braga

Entrega: até 16 de setembro de 2026 (aula 5). Peso: 10% da nota.

Esta lista deve ser entregue manuscrita. Escreva à mão, em papel, com as contas à vista — não só o resultado. Uma resposta certa sem desenvolvimento vale menos que uma resposta errada com o raciocínio visível, porque é no desenvolvimento que dá para ver onde está a dúvida.

Antes de começar

Esta lista **não é sobre ciência de dados**. Ela cobre álgebra linear, cálculo, estatística e probabilidade — as ferramentas que a disciplina usa o tempo todo e **não** ensina.

Os exercícios são deliberadamente básicos e diretos. Não há pegadinha, não há questão difícil escondida no fim. Se você resolver tudo sem sofrimento, ótimo: era esse o objetivo. Se travar em alguma parte, essa é a informação mais valiosa que esta lista produz — e ela chega na aula 5, quando ainda dá tempo de retomar.

Pode consultar livros, anotações e a internet. O que não pode é entregar uma resposta que você não saberia refazer sozinho, porque a prova vem em dezembro e ela não aceita consulta.

Cada parte abre dizendo **onde aquilo é usado no semestre**. Vale ler: a lista inteira foi montada a partir do que os capítulos realmente exigem.

1. Álgebra linear

Usada do Capítulo 4 em diante. Todo dado, neste curso, vira vetor: um usuário é um vetor de atributos, uma imagem é um vetor de pixels, e os parâmetros de um modelo são outro vetor. Produto escalar e norma aparecem em praticamente todos os algoritmos.

1. Sejam $v = [3, -1, 2]$ e $w = [1, 4, -2]$. Calcule:

- (a) $v + w$
- (b) $v - w$
- (c) $3v$

2. Com os mesmos v e w da questão anterior, calcule o produto escalar $v \cdot w$.

3. Calcule a norma (magnitude) de cada vetor:

- (a) $u = [3, 4]$
- (b) $v = [1, 2, 2]$

4. Calcule a distância euclidiana entre os pontos $p = (1, 2)$ e $q = (4, 6)$.

5. Uma matriz A tem 4 linhas e 3 colunas.

- (a) Qual a dimensão de A ?
- (b) O que representa o elemento $a_{2,3}$?
- (c) Quantos números a matriz guarda ao todo?

6. (conceitual) O produto escalar entre dois vetores não nulos deu **zero**. O que isso diz sobre eles?

2. Cálculo

Usado a partir do Capítulo 5, e depois em quase todo modelo. Treinar um modelo é minimizar uma função de erro, e o método para isso — gradiente descendente — é inteiramente construído sobre derivadas parciais.

7. Derive em relação a x :

- (a) $f(x) = x^3 - 5x + 2$
- (b) $g(x) = e^{2x}$
- (c) $h(x) = \ln(3x)$
- (d) $k(x) = \frac{1}{x}$

8. Use a regra da cadeia:

- (a) $f(x) = (2x + 1)^4$
- (b) $g(x) = e^{-x^2}$

9. Seja $f(x, y) = x^2y + 3xy^2 - y$. Calcule $\frac{\partial f}{\partial x}$ e $\frac{\partial f}{\partial y}$.

10. Seja $f(x, y) = x^2 + y^2$. Escreva o gradiente ∇f e calcule seu valor no ponto $(3, 4)$.

11. Encontre o valor de x que minimiza $f(x) = x^2 - 6x + 11$, e o valor mínimo de f .

12. A derivada pode ser estimada numericamente pelo quociente de diferenças:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

Para $f(x) = x^2$, estime $f'(3)$ usando $h = 0,01$. Compare com o valor exato.

13. (*conceitual*) Você está num ponto x e a derivada $f'(x)$ é **positiva**. Para **diminuir** o valor de f , você deve andar para a direita (aumentar x) ou para a esquerda (diminuir x)? Justifique em uma frase.

3. Estatística descritiva

Usada do Capítulo 7 em diante, e é o alicerce dos capítulos de regressão (11 e 12). Média e desvio padrão aparecem no reescalonamento de dados; covariância e correlação são a matéria-prima da regressão linear.

14. Considere o conjunto de dados: 2, 4, 4, 5, 10.

- (a) Calcule a média.
- (b) Calcule a mediana.
- (c) Calcule a moda.

15. Para o mesmo conjunto da questão anterior, calcule a **variância amostral** (divisor $n - 1$) e o desvio padrão.

16. Considere os pares (x, y) : $(1, 3)$, $(2, 5)$, $(3, 4)$, $(4, 8)$, $(5, 10)$.

- (a) Calcule a média de x e a de y .
- (b) Calcule a covariância amostral entre x e y .

17. (*conceitual*) Um conjunto de salários tem média R\$ 8.000 e mediana R\$ 3.000. O que isso sugere sobre a distribuição? Qual das duas medidas descreve melhor “o salário típico” nesse caso?

18. (*conceitual*) Em uma cidade, o número de sorvetes vendidos tem correlação alta com o número de afogamentos. Explique, em duas ou três frases, por que seria errado concluir que sorvete causa afogamento.

4. Probabilidade

Usada no Capítulo 10 (Naive Bayes) e no vocabulário de avaliação de modelos do Capítulo 8. Um classificador probabilístico é, literalmente, o teorema de Bayes aplicado a palavras.

19. Um dado honesto de seis faces é lançado uma vez. Calcule:

- (a) $P(\text{sair número par})$
- (b) $P(\text{sair número maior que 4})$

20. Uma urna tem 5 bolas vermelhas e 3 azuis. Duas bolas são retiradas, **sem reposição**. Sabendo que a primeira foi vermelha, qual a probabilidade de a segunda ser azul?

21. Uma moeda honesta é lançada duas vezes. Os dois lançamentos são eventos independentes.

- (a) Qual a probabilidade de sair cara nas duas vezes?
- (b) Escreva a fórmula geral de $P(A \cap B)$ quando A e B são independentes.

22. Uma doença atinge 1% da população. Um teste tem:

- sensibilidade de 99% (detecta a doença em 99% dos doentes);
- taxa de **falso positivo de 5%** (acusa positivo em 5% dos saudáveis).

Uma pessoa escolhida ao acaso testa **positivo**. Qual a probabilidade de ela estar realmente doente?

23. (*conceitual*) Explique, com um exemplo próprio, por que $P(A | B) \neq P(B | A)$ em geral.

5. Funções, logaritmos e somatórios

Usados nos Capítulos 13 (regressão logística), 14 (entropia) e 10 (Naive Bayes). A notação de somatório aparece em toda fórmula do curso.

24. Calcule:

- (a) $\log_2 8$
- (b) $\ln(e^3)$
- (c) $\log_{10} 100$

25. Escreva a propriedade que transforma o **logaritmo de um produto** numa soma, e explique em uma frase por que ela é útil ao multiplicar muitas probabilidades pequenas num computador.

26. A função logística é $\sigma(x) = \frac{1}{1+e^{-x}}$.

- (a) Calcule $\sigma(0)$.
- (b) Para onde tende $\sigma(x)$ quando $x \rightarrow +\infty$? E quando $x \rightarrow -\infty$?
- (c) (*conceitual*) Por que uma função com essa forma é conveniente para representar uma probabilidade?

27. Calcule $\sum_{i=1}^4 i^2$.

28. Escreva, usando notação de somatório, a fórmula da média de n valores $x_1, x_2, \dots, x_n$.

Antes de entregar

Confira três coisas:

1. **Está manuscrita?** Lista digitada não será aceita.

2. **As contas estão à vista?** Só o resultado final não basta.
3. **Você saberia refazer sozinho?** Se a resposta for não em alguma questão, marque-a e traga a dúvida para a aula — é para isso que esta lista existe.