

**Exercício 1.** Obtenha um classificador de Bayes ingênuo (Naive Bayes) no caso das flores iris. Obtenha métricas de desempenho considerando a divisão da base em treinamento (80%) e teste (20%). Você pode seguir os códigos disponíveis em

[https://github.com/cibelerusso/Aprendizado\\_de\\_Maquina](https://github.com/cibelerusso/Aprendizado_de_Maquina).

Varie o tamanho das bases de treinamento e teste e verifique o efeito em métricas de classificação.

**Exercício 2.** Obtenha um classificador por regressão logística para os dados amostra\_banco.csv, disponíveis em

<https://github.com/cibelerusso/IntroducaoInferenciaEstatistica/tree/main/Dados>.

Obtenha métricas de desempenho considerando a divisão da base em treinamento (80%) e teste (20%).

Você pode seguir os códigos disponíveis em

<https://github.com/cibelerusso/Modelos-de-Regressao>.

Varie o tamanho das bases de treinamento e teste e verifique o efeito em métricas de classificação.

**Exercício 3.** Considere um problema de classificação para um conjunto de dados

$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ , em que  $y_i \in C = \{c_1, c_2, \dots, c_k\}$ , utilizando a perda 0-1,  $L(y, \hat{g}(\mathbf{x})) = \mathbb{I}(Y \neq g(\mathbf{x}))$ .

Proponha um breve estudo de simulação para comparar os classificadores obtidos pelos métodos Naive Bayes e Regressão Logística, seguindo os passos:

- (1) Gere um conjunto de dados fictício com características  $(\mathbf{x}_i)$  e rótulos  $(y_i)$ . O conjunto de dados deve conter duas classes (binário), por exemplo, “spam” e “não spam”, ou “inadimplente” e “adimplente”. Você pode se inspirar em dados como iris ou amostra\_banco, que podem ser obtidos em [https://github.com/cibelerusso/Aprendizado\\_de\\_Maquina/tree/main/Dados](https://github.com/cibelerusso/Aprendizado_de_Maquina/tree/main/Dados) ou <https://github.com/cibelerusso/IntroducaoInferenciaEstatistica/tree/main/Dados> ou criar outros dados de sua preferência.
- (2) **Treinamento do Modelo** Divida o conjunto de dados em conjuntos de treinamento e teste (por exemplo, 80% para treinamento e 20% para teste).
- (3) Treine um modelo de Naive Bayes e um modelo de Regressão Logística no conjunto de treinamento.
- (4) **Avaliação do Desempenho:** Use os modelos treinados para fazer previsões no conjunto de teste.
- (5) Calcule a perda 0-1 para cada modelo em relação aos rótulos verdadeiros ( $L(y, \hat{g}(\mathbf{x})) = \mathbb{I}(Y \neq g(\mathbf{x}))$ ).
- (6) Calcule a taxa de erros (taxa de classificação incorreta) para cada modelo. Utilize outras métricas de desempenho.
- (7) **Repetição:** Repita as etapas 1 a 6 um número definido de vezes (por exemplo, 100 vezes) para obter médias e desvios padrão das taxas de erro para cada modelo.
- (8) **Análise dos Resultados:** Compare as médias das taxas de erro entre o modelo de Naive Bayes e o modelo de Regressão Logística.