**Ciência da Computação  
Instituto Federal do Ceará  
Campus Maracanaú**

**Professor: Hericson Araújo  
E-mail: hericson.araujo@ifce.edu.br  
Disciplina: Reconhecimento de Padrões**

**4ª Lista de Exercícios: Métodos Estatísticos e KNN**

**Instruções**

* **Com exceção dos casos explicitamente indicados, os algoritmos e modelos devem ser implementados do início em qualquer linguagem de programação (Python, Swift, Octave, etc.).**
* **Pacotes auxiliares (como sklearn, matplotlib, etc.) podem ser usados apenas para facilitar a manipulação de dados e a criação de gráficos.**
* **A entrega da solução pode ser feita via PDF ou Jupyter Notebook pelo Classroom. Os códigos também devem ser enviados.**

**Requisitos  
Desenvolver um relatório incluindo:**

1. **A parte teórica.**
2. **Equações matemáticas.**
3. **Imagens e metodologia utilizada.**

**Questões**

1. **Considere o conjunto de dados disponível em breastcancer.csv, organizado em 31 colunas:**
   * **As 30 primeiras colunas são os atributos.**
   * **A última coluna é a saída.**
   * **Os 30 atributos são coletados de exames médicos para diagnóstico do câncer de mama, com as seguintes classes:**
     + **1: Classe positiva.**
     + **0: Classe negativa.**
   * **Mais detalhes sobre os dados estão disponíveis em:**[**Documentação Scikit-Learn - Breast Cancer Dataset**](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html)**.**

**(a) Implemente os algoritmos:**

* **Análise de Discriminante Gaussiano.**
* **Naive Bayes Gaussiano.**
* **KNN (com k=3 e distância euclidiana).**

**(b) Usando validação cruzada com 10 folds, avalie os modelos implementados no item (a). Apresente:**

* **Valor médio e desvio padrão das métricas de:**
  + **Acurácia.**
  + **Revocação.**
  + **Precisão.**
  + **F1-score.**

**Observação:**

* **Os trabalhos devem ser implementados baseando-se nos slides da aula.**
* **Os cálculos de Média e Matriz de Covariância devem ser implementados pelo aluno.**

**Resumo dos Classificadores Estatísticos**

* **Análise de Discriminante Gaussiano**A fórmula é:  
  "y chapéu estrela é igual ao argumento que maximiza k da seguinte expressão:  
  log de p de Ck menos um meio vezes o log do determinante de Sigma k menos um meio vezes a transposta de (x estrela menos mu k) vezes a inversa de Sigma k vezes (x estrela menos mu k)."  
  Onde:
  + p(Ck) é a probabilidade a priori da classe Ck.
  + Sigma k é a matriz de covariância da classe Ck.
  + x estrela é o vetor de entrada a ser classificado.
  + mu k é o vetor de média da classe Ck.
* **Naive Bayes**A fórmula é:  
  "y chapéu estrela é igual ao argumento que maximiza k da seguinte expressão:  
  log de p de Ck mais a soma de d igual a 1 até D do log de p de x estrela d dado Ck."  
  Onde:
  + p(Ck) é a probabilidade a priori da classe Ck.
  + p(x estrela d | Ck) é a probabilidade condicional de cada atributo x estrela d dado a classe Ck.
  + D é o número de atributos (dimensão dos dados).
* **Naive Bayes Gaussiano**A fórmula é:  
  "y chapéu estrela é igual ao argumento que maximiza k da seguinte expressão:  
  log de p de Ck menos um meio vezes a soma de d igual a 1 até D do log de 2 vezes pi vezes sigma ao quadrado dk menos um meio vezes a soma de d igual a 1 até D de (x estrela d menos mu dk) ao quadrado dividido por sigma ao quadrado dk."  
  Onde:
  + p(Ck) é a probabilidade a priori da classe Ck.
  + D é o número de atributos (dimensão dos dados).
  + pi é a constante matemática pi, aproximadamente 3,1416.
  + sigma ao quadrado dk é a variância do atributo d na classe Ck.
  + mu dk é a média do atributo d na classe Ck.
  + x estrela d é o valor do atributo d do vetor de entrada x estrela.

**Observação:**Os parâmetros de todas as distribuições podem ser estimados a partir dos dados de treinamento disponíveis.