



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE SOBRAL
ENGENHARIA DA COMPUTAÇÃO E ENGENHARIA ELÉTRICA

TÓPICOS ESPECIAIS EM TELECOMUNICAÇÕES I (ECO0080)

TEMA: RECONHECIMENTO DE PADRÕES

Prof.: C. Alexandre Rolim Fernandes

Trabalho 4 – Classificador Bayesiano e LDA

- Trabalho Individual
- Sugere-se o uso do MATLAB ou Python para a prática de simulação.
- O código deve estar bem organizado e comentado, para que seja possível entendê-lo e corrigi-lo. Códigos que estejam desorganizados ou sem os devidos comentários explicativos terão penalização na nota.
- Fazer todas as questões em um só arquivo.
- O seu código deve gerar automaticamente todos os gráficos e resultados solicitados.
- Enviar no SIGAA apenas o arquivo do código (.m, .mlx ou .ipynb).
- O nome do arquivo deve ser igual ao seu nome.
- Não enviar código em formato PDF ou de texto (DOC, ODT, TXT etc).
- Não enviar arquivo compactado (zip,rar etc).
- Prazo e forma de entrega: dia 08/07/25 às 23h59, no SIGAA.

Prática de Simulação – Classificador Bayesiano e LDA com Dados Contínuos:

- Comece a prática gerando as duas seguintes matrizes:

$$\mathbf{A} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} c_5 & c_6 \\ c_7 & c_8 \end{bmatrix},$$

em que c_i é o i-ésimo número do seu CPF, para $i=1,2,\dots,8$.

- Em seguida, gere a duas matrizes \mathbf{X}_1 e \mathbf{X}_2 com dimensões 2×1000 e 2×2000 , respectivamente, contendo amostras de variáveis aleatórias Gaussianas com média 0 e variância igual a 1.
- Gere as matrizes: $\mathbf{Y}_1 = \mathbf{A} \mathbf{X}_1$ e $\mathbf{Y}_2 = \mathbf{B} \mathbf{X}_2$. Estas duas matrizes formam a base de dados inicial deste trabalho para as classes 1 e 2, respectivamente, em que $M=2$ representa o número de atributos, $N_1=1000$ representa no número de amostras da classe 1 e $N_2=2000$ representa no número de amostras da classe 2.
- A base de dados final é dada por $\mathbf{Z}_1 = \mathbf{Y}_1$ e $\mathbf{Z}_2 = \mathbf{Y}_2 + k \cdot \mathbf{M}_2$, em que \mathbf{M}_2 é uma matriz de dimensões 2×2000 contendo apenas o número 1 e k é um escalar. Você deve escolher o valor de k de forma que a maior dentre as acurácia obtidas nesta prática esteja acima de 96%. Em outras palavras, caso a acurácia de todas as técnicas testadas nesta prática estejam abaixo de 96%, você deverá deslocar (afastar) o centróide da nuvem de dados da classe 2 de forma a aumentar a acurácia, para que a melhor acurácia obtida seja maior do que 96%.
- Você deve implementar o classificador Bayesiano para classificar a base de dados acima, considerando os seguintes casos
 - Atributos Gaussianos - Caso geral;
 - Atributos Gaussianos e descorrelacionados (Naive Bayes);
 - Atributos Gaussianos, descorrelacionados e com variâncias iguais;
 - Atributos Gaussianos, descorrelacionados, com variâncias iguais e classes equiprováveis;
 - LDA seguido de um classificador unidimensional baseado em limiar (tal como visto em sala de aula).
- Para implementar estas técnicas, use as expressões do classificador Bayesiano que estão nos slides 45 e 46 do arquivo “Aula8.pdf”. Atenção, você deve se certificar que está com o arquivo mais atualizado do SIGAA.
- Usar K-fold com $K=10$.

- O algoritmo deve gerar as seguintes saídas:
 - Seu CPF
 - Gráfico de dispersão 2D dos dados das bases de dados inicial e final
 - Acurácia média dos 10 folds para cada classificador
 - Desvio-padrão da acurácia no K folds para cada classificador
 - Explicação, devidamente embasada, de porque o desempenho de cada técnica deu melhor ou pior (ou parecido) do que o das outras técnicas.
- Não há a etapa de extração de atributos. As entradas fornecidas já correspondem aos atributos.
- Não usar funções prontas para os classificadores Bayesinos nem para o LDA.