Universidade Federal do Ceará
Campus Sobral
Engenharia da Computação e Engenharia Elétrica

Tópicos Especiais em Telecomunicações I (ECO0080)
(Reconhecimento de Padrões)

Classificadores Bayesianos, Critério MAP e LDA

0) Informações Gerais

- Trabalho Individual. Simulação (código) e resposta às perguntas teóricas.

- Os códigos devem estar bem organizados e comentados, para que possa estar inteligíveis.

- Não usar "funções prontas" para cálculo de: Matriz de covariância; Matriz de correlação; Probabilidade à priori; matrizes de dispersão (Sw e Sb).

- Pode usar "funções prontas" para cálculo de: Vetor médio; Determinante; Inversa; Autovetores e Autovalores;

- Enviar as respostas e o código/implementação (Matlab / Python...) para o email:

david.coelho@sobral.ufc.br

- Prazo para entrega: 14/09/20 às 23:59.

1) Banco de dados:

1.1) Sistema de auxílio ao diagnóstico médico (Dermatologia)

- Diganóstico de doença de pele com base em informações clínicas (coletadas pelo médico no consultório) e informações hispatológicas (resultantes de uma biópsia – análise do tecido em um laboratório de patologia).

- 6 Classes (patologias): Psoríase (1), Dermatite Seborréica (2), Líquen plano (3), Pitiríase rósea (4), Dermatite Crônica (5), Pitiríase Rubrar Pilar (6).

- 34 atributos (figura 1).

- 358 amostras: 111 (classe 1) / 60 (classe 2) / 71 (classe 3) /

48 (classe 4) / 48 (classe 5) / 20 (classe 6) /

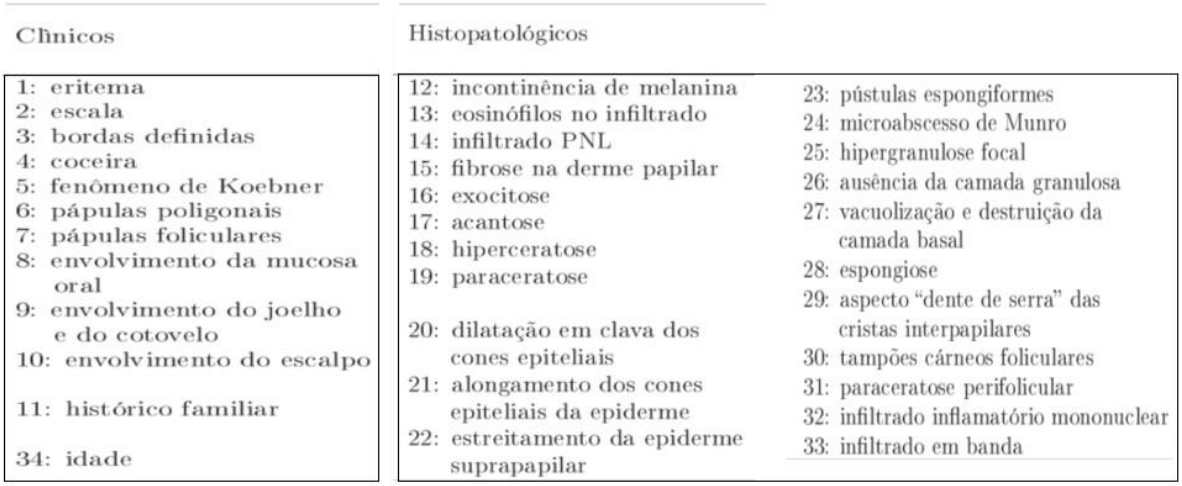

Figura 1. Atributos do banco de dados dermatologia.

2) Questões

2.1) Análise inicial dos dados

a) A partir do banco de dados completo (matriz **X** ∈ $\mathbb{R}^{Nxp}$ contendo todas os vetores de atributos), calcule:

- Vetor médio (média de cada atributo).

- Vetor de variâncias (variância de cada atributo).

- Matriz de covariância

- Matriz de correlação

b) A partir dos resultados obtidos no item anterior, responda:

- Que informação, sobre os atributos, é obtida através da matriz de correlação?

- Quais atributos são mais correlacionados?

2.2) Classificadores Bayesianos

- Implemente os seguintes classificadores bayesianos:

I) QDA (utilizando a formulação completa do discriminante bayesiano).

II) Naive Bayes (considerando que os atributos do problema são descorrelacionados)

III) LDA (considerando que todos os classificadores possuem a mesma matriz de covariância e a mesma probabilidade à priori).

- OBS: para calcular a matriz de covariância agregada, deve-se calcular as matrizes de covariância de cada classe, e calcular a média destas.

- Para cada classificador, utilizando o banco de dados “dermatologia” realize o seguinte experimento:

- Utilize a validação cruzada k-fold (k = 5) para gerar os resultados.
- Para cada um dos 5 particionamentos dos dados (entre treinamento e teste) realize o seguinte procedimento:
    - A partir dos dados de treinamento, calcule a média e o desvio padrão de cada atributo e faça a normalização z-score dos dados (tanto dos dados de treinamento como dos dados de teste).
    - Calcule as estatísticas necessárias (matriz de covariância, vetor médio, probabilidade a priori...), para o classificador, a partir dos dados de treinamento.
    - Classifique os dados de teste.

- Qual classificador obteve a melhor taxa de acerto (precisão, acurácia) média?

- Pode-se considerar esse problema linearmente separável? Justifique.

2.3) LDA (CDA) como transformação linear.

- A partir do banco de dados dermatologia, realize o seguinte experimento:

- Separe os dados entre treinamento (70%) e teste (30%).
- A partir dos dados de treinamento, calcule a média e o desvio padrão de cada atributo e faça a normalização z-score dos dados (tanto dos dados de treinamento como dos dados de teste).
- A partir dos dados de treinamento, aplique o método LDA utilizando a abordagem de múltiplas projeções (também conhecido como Canonical Discriminant Analysis, CDA).
- A partir dos autovalores calculados, defina uma quantidade de atributos necessária para realizar a projeção dos dados. (Obs: quantidade máxima = “número de classes – 1”)
- Defina a matriz de projeção W.
- Transforme os dados de treinamento a partir da matriz W.
- Utilize os dados transformados para treinar um classificador bayesiano (à sua escolha)
- Utilize a matriz de projeção transformar os dados de teste.
- Classifique os dados de teste e obtenha a taxa de acerto do classificador.