

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

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

Classificador Baseado em Árvores de Decisão

0) Informações Gerais

- Trabalho Individual. Apenas simulação, sem trabalho escrito.
- O código deve estar bem organizado e comentado, para que possa estar inteligível.
- Não usar “funções prontas” para o classificador.
- Enviar o código / implementação (Matlab / Python...) para o email:

david.coelho@sobral.ufc.br

- Prazo para entrega: 15/07/2021 às 23:59

1) Banco de dados: Wine

- 13 atributos contínuos. 3 classes (tipos de vinho – “1ª coluna do arquivo wine.csv”).
- 178 amostras: 59 da classe 1, 71 da classe 2 e 48 da classe 3.
- Os atributos já estão calculados, não precisa fazer a etapa de extração de atributos.

Componente	Mín	Máx	Média	DP
Álcool	11,0	14,8	13	0,8
Ácido málico	0,74	5,8	2,34	1,12
Cinza	1,36	3,23	2,36	0,27
Alcalinidade da cinza	10,6	30,0	19,5	3,3
Magnésio	70,0	162,0	99,7	14,3
Fenóis totais	0,98	3,88	2,29	0,63
Flavonoides	0,34	5,08	2,03	1,00
Fenóis não flavonoides	0,13	0,66	0,36	0,12
Proantocianidinas	0,41	3,58	1,59	0,57
Intensidade da cor	1,3	13,0	5,1	2,3
Matiz	0,48	1,71	0,96	0,23
OD280/OD315 de vinhos diluídos	1,27	4,00	2,61	0,71
Prolina	278	1680	746	315

2) Classificadores

2.1) Implementar o classificador: C4.5 (Árvore de Decisão)

- Deve usar a entropia (ganho de informação) como critério de escolha dos nós
- Não é necessário realizar as podas
- Você deve escolher os atributos que achar mais convenientes.
- Deve-se usar pelo menos 4 atributos.
- Não usar funções prontas do Matlab/Python para o C4.5

3) Experimento com o algoritmo C4.5 (fazer 10 realizações)

a) Realização:

- Dividir dados, aleatoriamente, entre treinamento e teste.
(Ex: 80% para treino, 20% para teste)
- Contrói o modelo com os dados de treinamento
- Verifica a acurácia (taxa de acerto) do modelo (teste).

b) Estatística:

- Calcular a taxa de acerto média das 10 realizações.