

# Aprendizagem de Máquina

## Lista de Exercícios 1

- 1) (2,5 pontos) Implemente os classificadores k-NN com e sem peso. Avalie os dois classificadores sobre duas bases de dados do UCI. Essas bases devem conter apenas dados numéricos. Varie o parâmetro  $k = \{1,2,3,5,7,9,11,13,15\}$  e construa um gráfico que mostre o comportamento da taxa de acerto à medida que o valor de  $k$  muda.
- 2) (2,5 pontos) Implemente a distância VDM. Escolha duas bases de dados do UCI (bases contendo apenas dados categóricos) e avalie a taxa de acerto dos dois classificadores implementados na Questão 1 (k-NN com e sem peso). Varie o número de vizinhos  $k = \{1,2,3,5,7,9,11,13,15\}$ .
- 3) (2,5 pontos) Implemente a distância para dados heterogêneos HVDM (Distância Euclidiana para atributos numéricos e distância VDM para atributos categóricos). Teste o k-NN com e sem peso sobre duas bases do UCI (escolha bases que contenham tanto atributos numéricos quanto categóricos). Varie o número de vizinhos  $k = \{1,2,3,5,7,9,11,13,15\}$ .
- 4) (2,5 pontos) Analise os resultados encontrados nas três questões anteriores quanto ao tempo de processamento e quanto à taxa de acerto, e construa uma argumentação que indique as melhores escolhas para as bases de dados avaliadas. (dica: observe o valor de  $k$ , o classificador com e sem peso, tempo de treinamento e tempo de teste).

Obs.:

- i) UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.html>
- ii) Cada aluno deve enviar por e-mail até a data estipulada os seguintes itens:
  1. Um relatório contendo a explicação detalhada por questão (metodologia dos experimentos, resultados, análise dos resultados);
  2. Código-fonte. Linguagens sugeridas: Matlab, Scilab, R project.
- iii) E-mail para [gdcc@cin.ufpe.br](mailto:gdcc@cin.ufpe.br) | Assunto da mensagem: [AM] 2015/01 – Lista1 (<seu login>)