

## UFPA PPGCC: Aprendizado de Máquina

Lista de Exercício Final - Valor 10 pts – Data de entrega 28/06/2019

- 1) [2.0 pts] Use os dados [Breast Cancer Wisconsin \(Diagnostic\) Data Set](#) do UCI Machine Learning Repository. Use validação cruzada para avaliar qual dos algoritmos tem maior acurácia nos dados :
- SVM Linear
  - SVM RBF

Decida que tipo de padronização (normalização) dos dados você usará para cada algoritmo (ou nenhuma). justifique.

2) [2.0 pts] Implemente em uma linguagem de programação de sua escolha uma Rede Neural Artificial Multilayer Perceptron treinada com o algoritmo backpropagation que resolva o problema do OU-EXCLUSIVO. Para validar sua implementação utilize a arquitetura apresentada na aula sobre Redes Neurais. Mostre que os resultados foram os mesmos. Avalie como ficaria a solução se considerarmos uma arquitetura com 2 neurônios na cada escondida e um neurônio na camada de saída. Lembrando que a função de ativação do neurônio deve ser sigmoide.

3) [2.0 pts] Dado o conjunto de dados abaixo:

a) aplique o método de agrupamento aglomerativo utilizando a métrica *single-link* e o critério de dissimilaridade distância Euclidiana.

b) aplique o algoritmo K-means utilizando distância Euclidiana considerando  $K = 3$ . O Algoritmo deve parar caso não apresente convergência após 5 iterações. Considere também que os centros iniciais são: cliente1 e cliente4.

c) Avalie qual melhor solução de clusterização considerando 3 grupos.

	X1	X2	X3	X4	X5
C_1	7,000	10,000	9,000	7,000	10,000
C_2	9,000	9,000	8,000	9,000	9,000
C_3	5,000	5,000	6,000	7,000	7,000
C_4	6,000	6,000	3,000	3,000	4,000
C_5	1,000	2,000	2,000	1,000	2,000
C_6	4,000	3,000	2,000	3,000	3,000
C_7	2,000	4,000	5,000	2,000	5,000

4) [2.0 pts] Implemente o método K-means. Os parâmetros de entrada são número K de clusters, o número M máximo de iterações, e um arquivo ARFF com o conjunto de treino (assuma que todos os atributos do ARFF devem ser levados em conta). O critério de

parada não precisa ser limitado a usar apenas o valor de  $M$ . Faça um tratamento (leve em conta) para o caso de algum cluster ficar com nenhum vetor associado a ele.

5) [2.0 pts] Fazer questão (slide 23 e 24) da aula sobre teste de significância do prof. Ronnie.