TIP8311 - Reconhecimento de Padrões

Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará (UFC)

Responsável: Prof. Guilherme de Alencar Barreto

2o. Trabalho Computacional - 14/11/2018

Questão Única - Tópico: Quantização vetorial para redução de grandes volumes de dados em problemas de classificação. Acesse através do link abaixo o conjunto de dados para uso neste trabalho.

`http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients`

Pede-se:

1. Compare os desempenhos dos classificadores Vizinho Mais Próximo (NN), Distância Mínima ao Centróide (DMC) e Classificador Quadrático Gaussiano (CQG), antes e depois da redução de volume. Preencha a tabela de resultados abaixo. Número de rodadas independentes de treino/teste: 100.

2. Qual dos classificadores se beneficiou mais da redução de volume?

Classif.	**Média**	**Mediana**	**[Mín./Máx.]**	**Desv. Pad.**	**Sensib.**	**Especif.**
NN						
DMC						
CQG						
NN-red						
DMC-red						
CQG-red						

Metodologia de Aplicação: Defina uma valor de K (e.g. $K = 1000$) e aplique o algoritmo K-médias aos dados de cada classe individualmente. Neste caso, cada classe terá $K = 1000$ protótipos, que substituirão os dados da respectiva classe.

Observação 1: Antes de aplicar os dados reduzidos aos classificadores, é importante rodar o K-médias um certo número de vezes (e.g. 10 vezes) a fim de escolher um posicionamento dos protótipos que produza o menor erro de quantização (medido pelo índice SSD).

Observação 2: Note que a técnica de redução de volume aqui discutida também equaliza ou equilibra o número de exemplos por classe, já que estamos usando o mesmo número de protótipos por classe.

Observação 3: Ainda sobre o usa da técnica de redução de volume para equalizar o número de amostras por classe, em muitos problemas de classificação uma das classes tem muito mais exemplos do que as demais. É o caso do conjunto de dados deste trabalho computacional. Nestes casos, a redução de volume por quantização vetorial pode ser aplicada apenas à classe com "excesso de dados".

Boa sorte!