

Universidade Federal do Ceará
Departamento de Engenharia de Teleinformática
Programa de Pós-Graduação em Engenharia de Teleinformática

TIP8311 – Reconhecimento de Padrões
Responsável: Prof. Dr. Guilherme de Alencar Barreto

5o. Trabalho Computacional – Semestre 2013-2

Data: 20/12/2013 - Data de entrega: 08/01/2014

Questão 1: Usando o conjunto de dados "DERMATOLOGY", aplicar o algoritmo de clusterização K -médias a este conjunto de dados sem a informação dos rótulos das classes. Determinar um valor adequado para K usando o índice de validação de clusters a seguir:

$$J_a(K) = \text{traço}(\mathbf{S}_w^{-1} \mathbf{S}_b)$$

em que \mathbf{S}_w é a matriz de espalhamento intra-clusters (*within-cluster scatter matrix*), \mathbf{S}_b é a matriz de espalhamento entre clusters (*between-cluster scatter matrix*) e o operador traço retorna a soma dos elementos da diagonal principal da matriz usada como argumento.

Para esta questão usar a seguinte sequência de passos:

Passo 1: Especificar K_{max} (no. máximo de clusters a serem avaliados) e R_{max} (no. máximo de repetições da execução do algoritmo K -médias para cada valor de K). Sugestão: $K_{max} = 15$ e $R_{max} = 30$.

Passo 2: Para $r = 1:R_{max}$, executar o algoritmo K -médias para o valor atual de K e calcular $J_a^{(r)}(K) = \text{traço}(\mathbf{S}_w^{-1} \mathbf{S}_b)$.

Passo 3: Calcular $\bar{J}_a(K) = \frac{1}{R_{max}} \sum_{r=1}^{R_{max}} J_a^{(r)}(K)$ (valor médio de $J_a^{(r)}(K)$ para um dado valor de K).

Passo 4: Fazer $K=K+1$.

Passo 5: Repetir os Passos 2 a 4, enquanto $K \leq K_{max}$.

Ao final, gerar o gráfico de $\bar{J}_a(K) \times K$. Escolher como valor adequado de K aquele que gerar o maior valor de $\bar{J}_a(K)$.

OBS: A função `kmeans` do Matlab pode ser usada para tratar esta questão.

Boa Sorte!