

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

4o. Trabalho Computacional – Semestre 2013-2

Data: 19/12/2013 - Data de entrega: 08/01/2014

Questão 1: Usando o conjunto de dados "DERMATOLOGY", aplicar PCA a este conjunto de dados para gerar novos conjuntos de vetores de dimensão reduzida Q . Para isso, resolver os seguintes itens:

1.1) Determinar um valor adequado para Q que conserve 95% da variância dos dados originais.

1.2) Determinar um valor adequado para Q que maximize a taxa de acerto para o novo conjunto de vetores. Usar a função `classify` do Matlab para implementar o classificador.

Os valores de Q obtidos nos itens anteriores são iguais?

OBS: As funções `pcacov` e `princomp` do Matlab podem ser usadas para resolver esta questão..

Questão 2: Usando o conjunto de dados "DERMATOLOGY", aplicar PCA para **selecionar** os atributos mais importantes/relevantes do conjunto de dados original. Determinar a taxa média de acerto resultante e seu desvio-padrão antes e após selecionar os atributos mais relevantes.

OBS: Atentar para a diferença entre **extração de atributos** (*feature extraction*) e **seleção de atributos** (*feature selection*). Na Questão 1, PCA está sendo usado para fins de **extração de atributos**, ou seja, os atributos originais estão sendo combinados linearmente para gerar os novos atributos.

Já na Questão 2, o que se pede é escolher/selecionar os atributos originais mais importantes usando PCA. Em outras palavras, escolher um certo número de atributos entre os 34 originais de tal forma que a taxa de acerto do classificador se mantenha aproximadamente a mesma.

Boa Sorte!