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 - 15/11/2023

Questão Única – Implementar/portar o código `demo_pca.m`, disponibilizado no SIGAA na linguagem de programação de sua preferência (por exemplo, Python, R, Julia ou outra).

Parte 1. Implementar a primeira parte do código que é a geração de dados artificiais com as características estatísticas desejadas de covariância. Use o método de Choleski para decompor a matriz de covariância desejada e correlacionar os 3 atributos gaussianos inicialmente não-correlacionados. Em suma, o objetivo é gerar a matriz `Xc`, com p=3 linhas e N=5000 vetores de atributos. Use a mesma matriz de covariância desejada Cd.

Parte 2. Implementar o passo-a-passo do algoritmo do PCA: (i) estimar a matriz de covariância Cx associada aos dados sintéticos originais na matrix Xc. (ii) Calcular os autovalores e a matriz de autovetores V. (iii) Comparar com os autovalores e autovetores calculados pelo código Matlab/Octave fornecido. (iv) Verificar se a matriz de autovetores é ortonormal de duas maneiras. A primeira é multiplicando ela por sua transposta. A outra é invertendo-a usando uma função de cálculo de matriz inversa (no Python) e comparando com a transposta de V. (v) Calcular a variância explicada por cada autovalor (VEi) e fazer o gráfico da variância explicada acumulativa (VEq).

Parte 3. Comparar os resultados do PCA implementado passo-a-passo na Parte 2 (método de livro-texto) com o método usando a decomposição em valores singulares (SVD, na sigla em Inglês).

Parte 4. Gerar os dados transformados; ou seja, gerar a matriz de dados Z. Verificar as propriedades esperadas para os dados transformados Z. Primeiro, numericamente, ao estimar a matriz de covariância Cz. Esta matriz é diagonal? Que elementos estão na diagonal principal de Cz? Em seguida, comparar o gráfico de dispersão (scatterplot) dos atributos X1 e X2 dos dados originais (correlacionados) e com o gráfico de Z1 e Z2 dos dados transformados (não correlacionados).

Parte 5. Projete os dados Xc em duas dimensões escolhendo apenas duas componentes (i.e. colunas da matriz de autovetores V). Em seguida, use a transposta da matriz Q para gerar a matriz de dados Xr, que é a matriz de reconstrução dos dados originais. (i) Mostre as 4 primeiras colunas de Xr e compare com as 4 primeiras colunas de Xc. Os valores são próximos? (ii) Calcule a norma quadrática de Frobenius da matriz de erro E= Xc – Xr. (iii) Vetorize a matriz E e calcule a soma dos erros quadráticos. Os valores de (i) e (ii) devem ser iguais.

Boa Sorte!