

Universidade Federal do Ceará - Campus de Quixadá

Prática 01

BCC0019 - Aprendizado de Máquina

Prof. Carlos Igor Ramos Bandeira

Outubro de 2016

Aluno (Nome/Matrícula): _____

A PONTUAÇÃO TOTAL QUE PODE SER CONQUISTADA COM ESTE TRABALHO É DE 10,0.

PRAZO DE ENTREGA: DEZEMBRO/2016.

1. (50 points) Usando o conjunto de dados “coluna.dat”, determinar:

- A matriz de covariância usando a função COV do Matlab/Octave. Observação: Realizar a operação de modo que os valores sejam divididos (normalizados) por “N” em vez de “N-1”, em que N denota o tamanho do conjunto de dados (i.e. número de observações).

Comparar o resultado obtido com as matrizes de covariância estimadas pelos seguintes métodos:

- **Método 1:** Usando a equação $\hat{\mathbf{C}}_x = \frac{1}{N} \sum_{i=1}^N [\mathbf{x}(i) - \bar{\mathbf{x}}][\mathbf{x}(i) - \bar{\mathbf{x}}]^T$, em que N é a quantidade de observações, $\mathbf{x}(i)$ é o vetor de dados no instante i e $\bar{\mathbf{x}}$ é o vetor de média.
- **Método 2:** Usando a equação $\hat{\mathbf{C}}_x = \hat{\mathbf{R}} - \bar{\mathbf{x}}\bar{\mathbf{x}}^T$. Observação: Para implementar o método 2, você pode optar por implementar a matriz de correlação $\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}(i) - \mathbf{x}^T(i)$, em que $\mathbf{x}(i)$ é o vetor de dados no instante i .
- **Método 3:** Usando a equação $\hat{\mathbf{R}}_x(n) = \frac{n-1}{n} \hat{\mathbf{R}}_x(n-1) + \frac{1}{n} \mathbf{x}(n)\mathbf{x}^T(n)$, em que n é a iteração atual, $\mathbf{x}(n)$ é o vetor sendo observado.

2. (25 points) Usando uma das matrizes de covariância estimadas na Questão 1, determinar os coeficientes de correlação entre todas as variáveis aleatórias envolvidas, usando os seguintes métodos:

- **Método 1:** Usando a função CORR do Matlab/Octave.
 - **Método 2:** Usando a função CORRCOEF do Matlab/Octave.
 - **Método 3:** Usando a equação $\rho_{ij} = \frac{c_{ij}}{\sigma_i \sigma_j}$, em que σ_i e σ_j são os desvios-padrões das variáveis i e j .
3. (25 points) Escolha 2 pares de variáveis (X_i, X_j) , sendo um par que apresente correlação negativa e um outro que apresente correlação positiva, e determine o gráfico de dispersão (scatterplot) para cada par. Coloque legendas nos eixos dos gráficos.