

Universidade Federal do Ceará
Departamento de Engenharia de Teleinformática
Curso de Engenharia de Computação

Projeto Computacional (TI0097 – Introdução ao Reconhecimento de Padrões)

Responsável: Prof. Guilherme de Alencar Barreto
Data de Início: 12/08/2020 - Data de entrega: 12/10/2020

Item 1 (Estatísticas Descritivas): Usando o conjunto de dados a você atribuído conforme a tabela mostrada na próxima página, pede-se:

1.1 - Escolher dois atributos quaisquer e calcular suas estatísticas descritivas (global e por classe), organizando-as conforme tabela abaixo.

| | Média | | Desvio-padrão | | Mediana | | Mínimo/Máximo | |
|-------------------------------|----------|--|---------------|--|----------|--|---------------|--|
| 1o. atributo escolhido | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |
| 2o. atributo escolhido | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |

1.2 – Escolha uma das classes (de preferência a que possuir mais dados) e estime o (i) seu centróide e (ii) sua matriz de covariância. Mostrar como foi realizado/feito/implementado este subitem.

1.3 – Para a classe do Subitem 1.2, fazer o gráfico de dispersão dos atributos escolhidos para resolução do Subitem 1.1. Colocar nomes dos atributos nos eixos dos gráficos.

1.4 – O diagrama de dispersão do Subitem 1.3 indica a existência de algum tipo de correlação entre os dois atributos escolhidos? Detalhe sua resposta.

1.5 - Estimar o coeficiente de correlação entre os dois atributos escolhidos para resolução do Subitem 1.3 para a classe escolhida. Mostrar os cálculos e qual método/procedimento foi utilizado para realizá-los.

1.6 – Determinar os parâmetros a (inclinação) e b (intercepto) da reta de tendência para os gráfico de dispersão mostrado no Subitem 1.3. Mostrar os cálculos e qual método/procedimento foi utilizado.

1.7 - Escolha um dos atributos usados na resolução do Subitem 1.1 e gere o histograma correspondente para uma classe de sua escolha. A distribuição do atributo escolhido assemelha-se a uma gaussiana? Há suspeita de presença de outliers nos dados? Detalhe sua resposta.

1.8 – Aplique o método de Box-Cox aos dados do Subitem 1.7 e, em seguida, aplique a transformação z-score (normalização para média zero e variância unitária). Forneça o histograma para o novo conjunto de medidas. A distribuição dos novos dados ficou mais próxima de uma gaussiana? Detalhe sua resposta.

Item 2 (Classificação de Padrões): Usando o mesmo conjunto de dados do Item 1, pede-se:

2.1 - Implementar classificadores estudados na disciplina e preencher a tabela abaixo.

| | Média | | Desvio-padrão | | Mediana | | Mínima/Máxima | |
|--|----------|--|---------------|--|----------|--|---------------|--|
| Vizinho mais Próximo | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |
| Classificador Linear de Mínimos Quadrados | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |
| Classificador Quadrático (Geral) | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |
| Classificador Quadrático (Variantes 1, 2 e 3) | Global | | Global | | Global | | Global | |
| | Classe 1 | | Classe 1 | | Classe 1 | | Classe 1 | |
| | ⋮ | | ⋮ | | ⋮ | | ⋮ | |
| | Classe K | | Classe K | | Classe K | | Classe K | |

Metodologia: Separar aleatoriamente 80% do conjunto de vetores para treinar o classificador linear dos mínimos quadrados e os 20% restantes serão usados para testar. Repetir 100 vezes o procedimento de treino-teste e determinar as estatísticas da tabela após as 100 rodadas.

2.2 – A utilização de PCA melhora o desempenho global de algum dos classificadores do Item 2.1? Em caso afirmativo, qual a dimensão dos vetores de atributos após a aplicação de PCA? Quanto da variância (i.e. informação) presente nos dados originais foi preservada?

Boa Sorte!

| Aluno | Conjunto |
|---|---|
| ALAN DE ABREU ESTEVAO | https://archive.ics.uci.edu/ml/datasets/Vertebral+Column |
| ANDRE FREIRE DE OLIVEIRA | https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra |
| DHANNYELL LUKAS BELMINO LIMA | https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records |
| DOMINGO SAVIO SILVA CARNEIRO | https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 |
| EDUARDO OLIVEIRA DE ALMEIDA | https://archive.ics.uci.edu/ml/datasets/Mushroom |
| FELIPE ARAUJO MAGALHAES | https://archive.ics.uci.edu/ml/datasets/Adult |
| FELIPE DO NASCIMENTO LOPES | https://archive.ics.uci.edu/ml/datasets/Energy+efficiency |
| JOSE MARCILIO DE SOUSA | https://archive.ics.uci.edu/ml/datasets/Wine+Quality |
| LUCA ISRAEL DE MOURA CRUZ | https://archive.ics.uci.edu/ml/datasets/South+German+Credit+%28UPDATE%29 |
| MIKAEL LUCAS DE BRITO SOUSA | https://archive.ics.uci.edu/ml/datasets/Wall-Following+Robot+Navigation+Data |
| RAUL COSTA FEITOSA | https://archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition |