

PRIMEIRA LISTA DE EXERCÍCIOS

HERBERTH AMARAL

Departamento de Ciência da Computação

Universidade Estadual de Montes Claros

herberthamaral@gmail.com

30 de agosto de 2015

Exercícios

1. 1 Objetivo

2. 2 Detalhamento

(a) 2.1

- i. a Número de exemplares: 8. Número de atributos: 5.
- ii. b Código de aluno: inteiro positivo. Nota 1 e nota 2: real. Hora de estudo: inteiro positivo. Situação: categórica (pode ser tratado como booleano também).
- iii. c Não há valores, atributos ou objetos ausentes. Também não há discrepâncias ou violações de domínio. Porém, como os valores das notas são próximos, podemos tentar diminuir a dimensão do dataset eliminando uma das notas. O atributo de código de aluno também é descartável. As horas de estudo a primeira vista também parecem ser um indicador de desempenho da nota. A situação do aluno pode ser descartada porque é obtida com uma regra básica ($\text{nota} > 6$).
- iv. d Conforme algoritmos implementados e disponíveis no arquivo *lista1.py*, a correlação das notas utilizando o método de Pearson é de 0.999844482844, o que torna uma das notas um candidato a exclusão. Já a correlação da normalização utilizando z-score de uma das notas e das horas de estudo é 0.695484993417, que chega bem próximo, mas não atinge o limiar proposto de > 0.7 ou < -0.7 .

(b) 2.2

- i. a Número de exemplares: 12. Número de atributos: 4. Com relação aos tipos: altura e largura são inteiros, aspecto é real, e classe é categórica/booleana.
- ii. b É necessário alguns passos de pré-processamento. Primeiramente há valores faltantes. Alguns desses valores podem ser inferidos utilizando o fato que o aspecto é a razão entre a largura e a altura. Obviamente a inferência não será possível quando mais de um valor dentre a altura, largura e aspecto estiver faltando. Neste caso, pode-se considerar a imputação desses valores. Pelos cálculos feitos e disponíveis no arquivo *lista1.py*, não há grandes correlações entre variáveis normalizadas.
- iii. c Disponível no arquivo *lista1.py*.