



## **PRÁTICA 2**

Deseja-se construir um aplicativo que receba os dados digitados referentes a exemplos da espécie de lagarto "*Cophosaurus texanus*" e que indique se este exemplo é do sexo feminino ou masculino. Para o treinamento e teste do classificador, será utilizado o conjunto de dados anexo.

Para muitos classificadores, é importante normalizar o conjunto de treinamento. Desejamos normalizar os valores de cada atributo individualmente. Uma maneira de fazer isso é achar os mínimos e máximos de cada coluna de atributos e utilizar estes valores para normalizar os dados. Considere que desejamos normalizar os dados de cada coluna para valores entre -1 e +1. Considere que os conjuntos de treinamento e teste são formados por conjuntos disjuntos com, respectivamente,  $p \cdot N$  e  $(1-p) \cdot N$  exemplos do conjunto de dados, sendo  $N$  o tamanho do conjunto de dados original e  $p$  um número entre 0 e 1 que controla o tamanho dos conjuntos de treinamento e teste. Pede-se:

1. Dado um valor de  $p$ , gere os conjuntos de treinamento e teste com aproximadamente a mesma proporção de dados de cada uma das classes (ou seja, os balanceamentos dos conjuntos de treinamento e teste devem ser similares).
2. Apresente a fórmula para normalizar os valores do conjunto de treinamento. Mostre como chegou nesta fórmula.
3. Normalize o conjunto de treinamento para  $p=0,75$ .
4. Gere um programa que receba os dados (atributos) para um exemplo qualquer do lagarto e mostre os dados normalizados (utilizando a mesma normalização utilizada para normalizar o conjunto de treinamento, i.e., utilizando a fórmula de normalização com os máximos e mínimos encontrados no conjunto de treinamento).