

PRÁTICA 2

Larissa Magalhães Pereira
13747904

1. O programa faz a leitura do arquivo que contém todos os dados, armazenando as informações dos sexos dos lagartos em um vetor `sexo[]`. Posteriormente, ele pede ao usuário para informar um valor `p` (um número entre 0 e 1 que controla o tamanho dos conjuntos de treinamento e teste) e, com base nesse valor, separa os dados em conjuntos de treinamento e teste seguindo uma proporção igual entre as classes (feminino e masculino). Por exemplo, com $p = 0.75$ e sendo 13 machos e 12 fêmeas, temos que no conjunto treinamento terão $0.75 * 13 \approx 10$ machos e $0.75 * 12 = 9$ fêmeas, com proporção $10/9 = 1,11$. Já no conjunto teste, terão $0.25 * 13 \approx 3$ machos e $0.25 * 12 = 3$ fêmeas, com proporção $3/3 = 1$.

```
Informe um valor para p (valor entre 0 e 1 - controla o tamanho dos conjuntos de treinamento e teste): 0.75

Q1. Gerando os conjuntos treinamento e teste com aproximadamente a mesma proporcao de dados de cada uma das classes, com base em p
Conjunto Treinamento:
1: 5.52600, 59.00000, 113.50000, F
2: 10.40100, 75.00000, 142.00000, M
3: 9.21300, 69.00000, 124.00000, F
4: 8.95300, 67.50000, 125.00000, F
5: 7.06300, 62.00000, 129.50000, M
6: 6.61000, 62.00000, 123.00000, F
7: 11.27300, 74.00000, 140.00000, M
8: 2.44700, 47.00000, 97.00000, F
9: 15.49300, 86.50000, 162.00000, M
10: 9.00400, 69.00000, 126.50000, F
11: 8.19900, 70.50000, 136.00000, M
12: 6.60100, 64.50000, 116.00000, F
13: 7.62200, 67.50000, 135.00000, M
14: 10.06700, 73.00000, 136.50000, M
15: 10.09100, 73.00000, 135.50000, M
16: 10.88800, 77.00000, 139.00000, M
17: 7.61000, 61.50000, 118.00000, F
18: 7.73300, 66.50000, 133.50000, M
21: 5.14900, 59.50000, 116.00000, F

Conjunto Teste:
19: 12.01500, 79.50000, 150.00000, M
20: 10.04900, 74.00000, 137.00000, M
22: 9.15800, 68.00000, 123.00000, F
23: 12.13200, 75.00000, 141.00000, M
24: 6.97800, 66.50000, 117.00000, F
25: 6.89000, 63.00000, 117.00000, F
```

2. A fórmula usada para normalizar os valores dos conjuntos de treinamento foi:

$$\left(\frac{v - \min}{\max - \min} \right) * 2 - 1$$

- v = valor a se normalizar
- \min = valor mínimo da coluna
- \max = valor máximo da coluna

Vamos supor que se $v = \min$, o resultado é 0. E, se $v = \max$, o resultado é 1. Com base nisso, podemos deduzir a equação $\left(\frac{v - \min}{\max - \min} \right)$. Porém, queremos chegar em -1 para o valor mínimo e 1 para o valor máximo, assim, basta multiplicarmos essa equação por 2 e depois subtrair 1.

3. Normalizando o conjunto de treinamento para $p = 0.75$, temos:

Q3. Conjunto de treinamento para $p = 0.75$ normalizado:

```
1: -0.52798, -0.39241, -0.49231, F
2: 0.21938, 0.41772, 0.38462, M
3: 0.03725, 0.11392, -0.16923, F
4: -0.00261, 0.03797, -0.13846, F
5: -0.29235, -0.24051, 0.00000, M
6: -0.36180, -0.24051, -0.20000, F
7: 0.35306, 0.36709, 0.32308, M
8: -1.00000, -1.00000, -1.00000, F
9: 1.00000, 1.00000, 1.00000, M
10: 0.00521, 0.11392, -0.09231, F
11: -0.11820, 0.18987, 0.20000, M
12: -0.36318, -0.11392, -0.41538, F
13: -0.20665, 0.03797, 0.16923, M
14: 0.16817, 0.31646, 0.21538, M
15: 0.17185, 0.31646, 0.18462, M
16: 0.29404, 0.51899, 0.29231, M
17: -0.20849, -0.26582, -0.35385, F
18: -0.18964, -0.01266, 0.12308, M
```

O programa percorre 75% dos dados totais (conjunto de treinamento) por meio dos vetores `massa[]`, `comprimento[]` e `dimensao[]`, normalizando cada um de seus valores a partir da fórmula $\left(\frac{v - \min}{\max - \min}\right) * 2 - 1$, sendo o max e o min respectivos a cada vetor/coluna.

4. Normalizar para um exemplo qualquer (com base no p dado anteriormente):

Q4. Normalizando dados de um exemplo qualquer de um lagarto

Informe os dados do lagarto:

Massa: 7

Comprimento: 68

Dimensao: 120

Sexo (M ou F): F

Normalizado: -0.30201, 0.06329, -0.29231, F

O programa pede o usuário para informar os dados (massa, comprimento, dimensão e sexo) do lagarto e aplica a fórmula $((\text{valor} - \min)/(\max - \min) * 2 - 1)$ para cada variável, com base no conjunto de treinamento obtido anteriormente (em p e nos máximos e mínimos), obtendo a normalização dos dados.