

**Trabalho Final da Disciplina de Tópicos Especiais da Computação**  
**Gláuber Barbacovi Ribeiro - 103179**  
**Turma Ciência da Computação - 2022/2**

**Conjunto de Dados - Abalone**

O conjunto de dados Abalone é usado com frequência para problemas de regressão, pois o objetivo principal é prever a idade de um abalone (um tipo de molusco marinho) com base em suas características físicas. Esse dataset contém informações como comprimento, diâmetro, peso, e o número de anéis, que é usado para estimar a idade.

A idade do abalone é determinada cortando a concha através do cone, colorindo-a e contando o número de anéis através de um microscópio -- uma tarefa chata e demorada. Outras medições, que são mais fáceis de obter, são usadas para prever a idade. Mais informações, como padrões climáticos e localização (portanto, disponibilidade de alimentos) podem ser necessárias para resolver o problema. Dos dados originais, exemplos com valores ausentes foram removidos (a maioria com o valor previsto ausente), e os intervalos dos valores contínuos foram dimensionados para uso com uma ANN (dividindo por 200).

**Tabela de Variáveis**

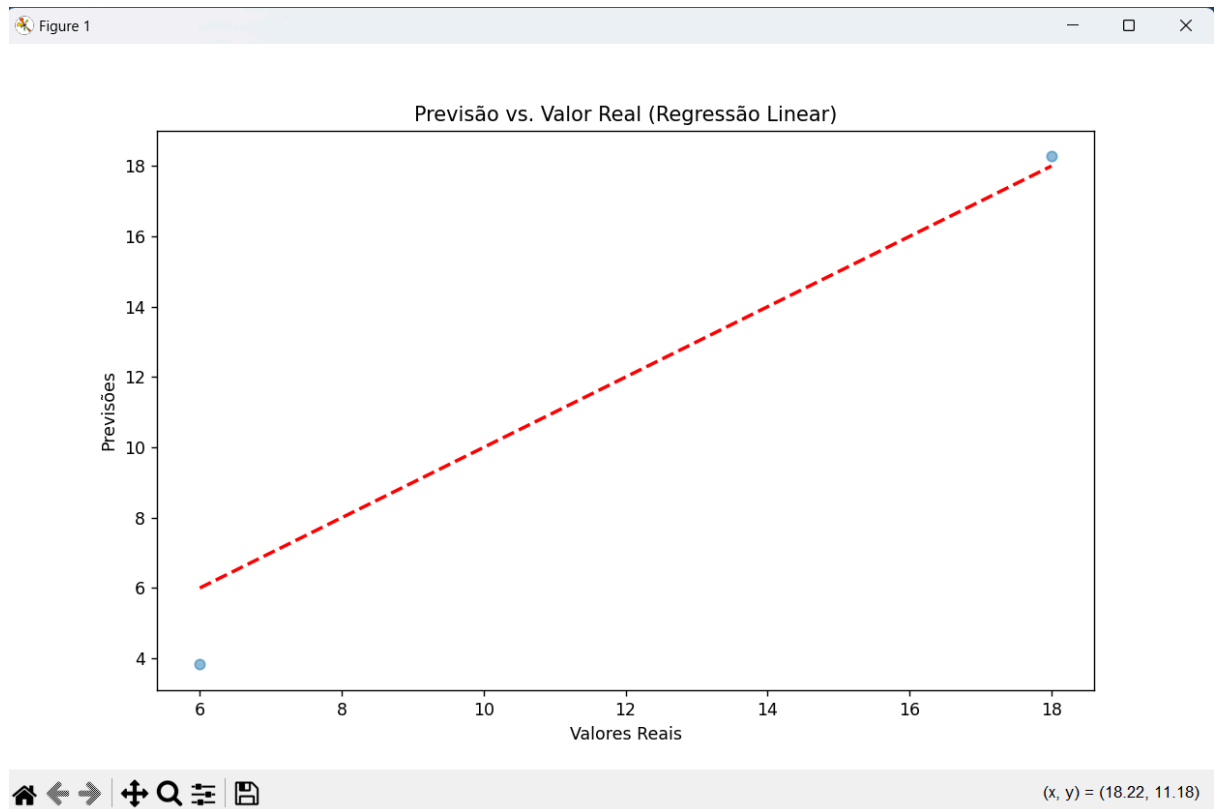
Nome da Variável	Papel	Tipo	Descrição	Unidades	Valores ausentes
Sexo	Recurso	Catégorico	M, F e I (bebê)		não
Comprimento	Recurso	Contínuo	Medida da concha mais longa	milímetros	não
Diâmetro	Recurso	Contínuo	perpendicular ao comprimento	milímetros	não
Altura	Recurso	Contínuo	com carne com casca	milímetros	não
Peso_inteiro	Recurso	Contínuo	abalone inteiro	gramas	não
Peso_descascado	Recurso	Contínuo	peso da carne	gramas	não
Peso das vísceras	Recurso	Contínuo	peso intestinal (após sangramento)	gramas	não
Peso da concha	Recurso	Contínuo	depois de ser seco	gramas	não
Anéis	Alvo	Inteiro	+1,5 dá a idade em anos		não

## Análise e Aplicação

Com o dataset carregado, foi realizado teste da técnica de regressão com as seguintes informações:

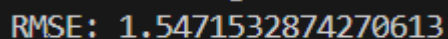
```
def load_data():  
    # Dados fictícios (expandido para mais moluscos)  
    X = np.array([  
        [0.5, 0.35, 0.1, 0.2, 0.3, 0.15, 0.07], # Molusco 1  
        [0.6, 0.4, 0.2, 0.25, 0.4, 0.2, 0.08], # Molusco 2  
        [0.7, 0.5, 0.15, 0.3, 0.35, 0.18, 0.09], # Molusco 3  
        [0.8, 0.6, 0.25, 0.4, 0.5, 0.3, 0.1], # Molusco 4  
        [0.9, 0.7, 0.2, 0.5, 0.6, 0.4, 0.12], # Molusco 5  
        [1.0, 0.75, 0.3, 0.6, 0.65, 0.45, 0.15], # Molusco 6  
        [0.85, 0.65, 0.35, 0.55, 0.55, 0.4, 0.13], # Molusco 7  
        [0.95, 0.8, 0.28, 0.65, 0.7, 0.5, 0.14], # Molusco 8  
        [1.1, 0.85, 0.4, 0.75, 0.8, 0.6, 0.2], # Molusco 9  
        [1.2, 0.9, 0.45, 0.85, 0.9, 0.7, 0.25] # Molusco 10  
    ])  
  
    # Idades dos moluscos (alvo)  
    y = np.array([3, 6, 8, 10, 12, 15, 14, 16, 18, 20]) # Idades fictícias  
    # Metadados
```

Aqui é possível ver as características dos 10 abalones. Foi escolhido uma idade para cada um que “é a idade correta” e com as informações de cada abalone corretas o programa irá encontrar a idade deles que está prevista.



O gráfico gerado mostra que o programa tem uma margem de erro baixíssima uma vez que previu que a idade deles começa aproximadamente em 6 anos (esperado era 3 anos) e finaliza em quase 18 anos (esperado era 20 anos).

O índice de erro reduz na medida que aumenta a quantidade de animais testados por vez, temos por fim o RMSE Root Mean Squared Error (RMSE), ou Erro Quadrático Médio é esse:



```
RMSE: 1.5471532874270613
```

Portanto é possível afirmar que as idades estão em torno de 4,5 anos até 19,5 anos, assim satisfazendo o esperado.

Foram realizados testes com apenas 3 exemplos de abalone e o RMSE foi mais alto chegando a 2.57 anos de diferença.

## Referências Bibliográficas

**Wikipedia.** ABALONE. Disponível em: <https://pt.wikipedia.org/wiki/Abalone>. Acesso em 15 nov. 2024

**Semantic Scholar.** CUSTODES: AUDITABLE HYPOTHESIS TESTING. Disponível em: <https://www.semanticscholar.org/paper/Custodes%3A-Auditable-Hypothesis-Testing-Servan-Schreiber-Ohrimenko/d40d9ed8979aa7d6be9c5cfd31db3936c5830839>. Acesso em: 15 nov. 2024

**UC Irvine.** ABALONE. Disponível em: <https://archive.ics.uci.edu/dataset/1/abalone>. Acesso em 15 nov. 2024