Atividade Avaliativa de Álgebra linear

Lucas Emanoel, Rafael Emanuel, Éricles Barros
João Pedro Ferreira, Ruslan Andruscha

25 de março de 2024

1 Escolhendo uma base de dados

Nossa equipe optou por utilizar o Datasheet nº2 que contém os seguintes dados:

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 1	52	158
Pessoa 2	92	175
Pessoa 3	55kg	172cm
Pessoa 4	54	155
Pessoa 5	80	174
Pessoa 6	45kg	150cm
Pessoa 7	68	167
Pessoa 8	70	1.68
Pessoa 9	68	167
Pessoa 10	62	175

Tabela 1: Dados de peso e altura das pessoas.

Dividimos esses dados em 2 conjuntos. O primeiro, chamado Conjunto de Treino, contém as primeiras 7 instâncias da nossa base de dados.

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 1	52	158
Pessoa 2	92	175
Pessoa 3	55kg	172cm
Pessoa 4	54	155
Pessoa 5	80	174
Pessoa 6	45kg	150cm
Pessoa 7	68	167

Tabela 2: Conjunto de Treino.

O segundo conjunto foi nomeado como Conjunto de Teste e compreende as últimas 3 instâncias da nossa base de dados:

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 8	70	1.68
Pessoa 9	68	167
Pessoa 10	62	175

Tabela 3: Conjunto de Teste.

2 Tratando a base de dados

Visando igualar a natureza dos dados que vamos manipular, trataremos o peso e altura como valores inteiros, eliminando informações irrelevantes e padronizando suas unidades. A nova base de dados terá a seguinte forma:

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 1	52	158
Pessoa 2	92	175
Pessoa 3	55	172
Pessoa 4	54	155
Pessoa 5	80	174
Pessoa 6	45	150
Pessoa 7	68	167
Pessoa 8	70	168
Pessoa 9	68	167
Pessoa 10	62	175

Tabela 4: Dados de peso e altura após o tratamento

Essa simples manipulação é extremamente poderosa e permite que possamos escrever o Conjunto de treino como sendo:

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 1	52a	158
Pessoa 2	92a	175
Pessoa 3	55a	172
Pessoa 4	54a	155
Pessoa 5	80a	174
Pessoa 6	45a	150
Pessoa 7	68a	167

Tabela 5: Conjunto de Treino.

Como veremos a seguir, esse formato será muito útil e posteriormente permitirá a formação de um sistema linear preditivo.

3 Obtendo um sistema linear

Se considerarmos a coluna de pesos como valores de x e a coluna das alturas como valores de y, podemos obter o seguinte sistema linear:

1. 52a + b = 158
2. 92a + b = 175
3. 55a + b = 172
4. 54a + b = 155
5. 80a + b = 174
6. 45a + b = 150
7. 68a + b = 167

4 Sistema no Formato Matricial

Podemos reescrever o sistema linear anterior no formato matricial. Assim, nosso sistema adquire a forma:

$$\begin{bmatrix} 52 & 1 \\ 92 & 1 \\ 55 & 1 \\ 54 & 1 \\ 80 & 1 \\ 45 & 1 \\ 68 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 158 \\ 175 \\ 172 \\ 155 \\ 174 \\ 150 \\ 167 \end{bmatrix}$$

Onde a primeira matriz é chamada de matriz dos coeficientes, a segunda de matriz das incógnitas e a terceira de matriz dos termos independentes.

5 Atribuindo Nomes e Tratando o Sistema

Adotaremos A para referir-se à matriz dos coeficientes, X para referir-se à matriz das incógnitas e Y para referir-se à matriz dos termos independentes. Com isso, temos que:

$$A = \begin{bmatrix} 52 & 1 \\ 92 & 1 \\ 55 & 1 \\ 54 & 1 \\ 80 & 1 \\ 45 & 1 \\ 68 & 1 \end{bmatrix} \qquad X = \begin{bmatrix} a \\ b \end{bmatrix} \qquad Y = \begin{bmatrix} 158 \\ 175 \\ 172 \\ 155 \\ 174 \\ 150 \\ 167 \end{bmatrix}$$

Queremos $A^\top A$ e $A^\top Y$. Para tal, calcularemos primeiramente $A^\top$.

$$A^\top = \begin{bmatrix} 52 & 92 & 55 & 54 & 80 & 45 & 68 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Logo, $A^\top A$ trata-se da matriz:

$$A^\top A = \begin{bmatrix} 52 & 92 & 55 & 54 & 80 & 45 & 68 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 52 & 1 \\ 92 & 1 \\ 55 & 1 \\ 54 & 1 \\ 80 & 1 \\ 45 & 1 \\ 68 & 1 \end{bmatrix} = \begin{bmatrix} 30158 & 446 \\ 446 & 1 \end{bmatrix}$$

Podemos também calcular $A^\top Y$:

$$A^\top Y = \begin{bmatrix} 52 & 92 & 55 & 54 & 80 & 45 & 68 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 158 \\ 175 \\ 172 \\ 155 \\ 174 \\ 150 \\ 167 \end{bmatrix} = \begin{bmatrix} 74172 \\ 1151 \end{bmatrix}$$

6 Encontrando a Solução do SL $A^\top AX = A^\top Y$

$$\begin{bmatrix} 30158 & 446 \\ 446 & 1 \end{bmatrix} \cdot \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} 74172 \\ 1151 \end{bmatrix}$$

$$30158a + 446b = 74172 \tag{1}$$
$$446a + b = 1151 \tag{2}$$

A partir de (2) temos:

$$b = 1151 - 446a \tag{3}$$

Substituindo(3) em (1), temos:

$$\begin{aligned} &30158a + 446(1151 - 446a) = 74172 \\ &30158a + 513346 - 198916a = 74172 \\ &-168758a = 74172 - 513346 \\ &168758a = 439174 \end{aligned}$$

$$a \approx 2,6 \tag{4}$$

Obtendo o valor de a podemos substituir (4) em (2):

$$446a + b = 1151$$
$$1159,6 + b = 1151$$

$$b = -8,6 \tag{5}$$

7 Encontrando a Equação da Reta

No item anterior, descobrimos os valores dos coeficientes a e b

$$a \approx 2,6$$
$$b = -8,6$$

Temos que a forma geral da equação de uma reta é dado por

$$y = ax + b \tag{6}$$

Substituindo (3) e (4) em (6) é possível encontrar a equação da reta que melhor descreve o comportamento do nosso sistema linear:

$$r : 2,6x - 8,6 \tag{7}$$

A próxima etapa consiste em utilizar o Conjunto de Teste para quantificar a precisão do nosso modelo.

Pessoa	**Peso (kg)**	**Altura (cm)**
Pessoa 8	70	168
Pessoa 9	68	167
Pessoa 10	62	175

Tabela 6: Conjunto de Teste.

8 Teste do Modelo Linear

Utilizando os dados contidos no Conjunto de Teste, é possível realizar previsões baseadas no comportamento da equação linear encontrada no item anterior.

Tais previsões podem não apresentar o real valor da altura do indivíduo, mas deve representar uma boa aproximação desde que o erro relativo esteja contido no intervalo esperado.

É possível calcular o erro relativo utilizando a seguinte expressão:

$$Erro_{\text{relativo}} = \frac{Medida_{\text{real}} - Medida_{\text{prevista}}}{Medida_{\text{real}}} \times 100\%$$

Em posse dessa equação e com um modelo de previsão, foi possível gerar os seguintes dados:

Nome	Peso (kg)	Altura (cm)	Prev. do Modelo (cm)	Erro Relativo
Pessoa 8	70	168	173,4	3,21%
Pessoa 9	68	167	168,2	0,84%
Pessoa 10	62	175	152,6	12,80%

Tabela 7: Previsão do Modelo

9 Enriquecendo nosso Modelo

O objetivo desse item é analisar a correlação entre os dados que estamos estudando, além de finalmente adentrar no âmbito computacional e de fato analisar o funcionamento do modelo.

Dentre as tecnologias aptas a desenvolver essa atividade, a equipe optou por utilizar-se da linguagem de programação Python. Utilizamos a biblioteca pandas para manipular a base de dados, matplotlib para plotar o gráfico de dispersão e numpy para calcular o coeficiente de correlação entre os dados do nosso sistema. O código gerado pode ser encontrado no repositório público **atividadeAlgebra**.

O coeficiente de correlação ρ pode assumir valores entre -1 e 1, onde:

$$\rho = 0,9 \text{ a } 1 \quad \text{(positivo ou negativo) : correlação muito forte;}$$
$$\rho = 0,7 \text{ a } 0,9 \quad \text{(positivo ou negativo) : correlação forte;}$$
$$\rho = 0,5 \text{ a } 0,7 \quad \text{(positivo ou negativo) : correlação moderada;}$$
$$\rho = 0,3 \text{ a } 0,5 \quad \text{(positivo ou negativo) : correlação fraca;}$$
$$\rho = 0 \text{ a } 0,3 \quad \text{(positivo ou negativo) : não possui correlação.}$$

Utilizando o método corrcoef contido na biblioteca numpy, foi possível verificar que o ρ da nossa base de dados é dado por:

$$\rho = 0,742$$

Esse valor é expressivo e indica que os nossos dados possuem uma forte correlação entre si.

O próximo passo foi plotar o gráfico de dispersão dos dados. Para isso, foram utilizadas diversas ferramentas da biblioteca matplotlib. O resultado pode ser visualizado abaixo:

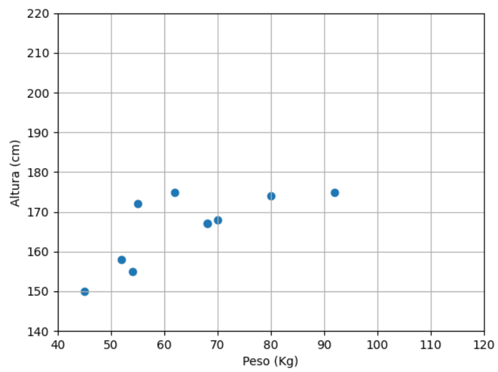

Figura 1: Gráfico de dispersão do Datasheet 2

Verificamos também que essa seria uma excelente oportunidade para testar o modelo de previsão que vínhamos desenvolvendo durante toda a atividade. Adicionalmente ao gráfico de dispersão, plotamos a equação (7), obtendo o gráfico à seguir:

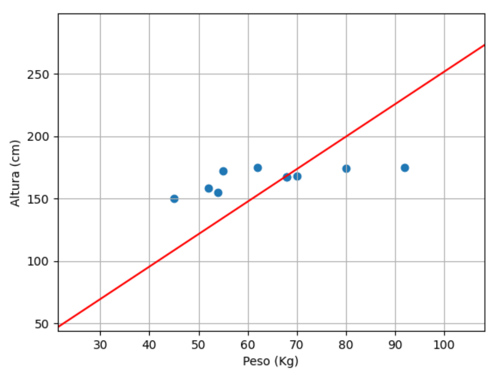

Figura 2: Reta r: 2,6x - 8,6 aplicada ao gráfico de dispersão

Com isso, podemos concluir que o nosso modelo apresenta uma aproximação satisfatória dos resultados reais esperados.

10 Considerações Finais

Após completar a atividade avaliativa de Álgebra Linear, nossa equipe alcançou diversas conclusões significativas. Primeiramente, ficou claro como os conceitos de Álgebra Linear são fundamentais e aplicáveis em contextos do mundo real, como a modelagem de dados de peso e altura. Através da construção de um sistema linear e da resolução matricial, pudemos compreender como esses conceitos são essenciais na formulação e resolução de problemas complexos.

Além disso, a atividade nos permitiu aplicar esses conceitos utilizando ferramentas computacionais modernas, como Python, o que reforçou a importância da programação na análise e interpretação de dados. Através da manipulação e visualização dos dados, pudemos entender melhor o relacionamento entre as variáveis e avaliar a qualidade do modelo linear proposto.