SIMPÓSIO DE INTEGRAÇÃO, INOVAÇÃO E TECNOLOGIA

ABORDAGEM MATRICIAL PARA O MODELO DE REGRESSÃO LINEAR

Fabio Nogueira Carlucci[1]

RESUMO: Este trabalho está focado em um dos papeis da inteligência artificial (IA) na álgebra linear. Mais precisamente, quais as implicações no ajuste de uma função do primeiro grau, aos dados observados, obtido pelo método dos mínimos quadrados ordinários por regressão linear escalar ou matricial. Nesse sentido, foi realizado um estudo sobre um novo algoritmo para a multiplicação de matrizes, utilizando o método de Strassen e foram citados os métodos de multiplicação que utilizam a Inteligência Artificial (IA) como o AlphaTensor, da DeepMind, que acarretam diminuição no tempo de processamento para grandes bancos de dados e, consequentemente, maior economia dos fatores de produção. Ao final foi realizado um produto de matrizes pelo método tradicional e pelo método de Strassen, sendo constatado que esse método é melhor do que o método tradicional de Cayley, quando o parâmetro é o tempo de processamento.
Palavras-chave: matrizes; inteligência artificial; tempo.

ABSTRACT: This paper focuses on one of the roles of artificial intelligence (AI) in linear algebra. More specifically, the implications of fitting a first-degree function to observed data obtained by the ordinary least squares method by scalar or matrix linear regression. In this regard, a study was conducted on a new algorithm for matrix multiplication using the Strassen method. We also reported on multiplication methods that use Artificial Intelligence (AI), such as DeepMind's AlphaTensor, which result in reduced processing time for large databases and, consequently, greater savings in production factors. Finally, a matrix product was performed using the traditional method and the Strassen method, and it was found that this method outperforms the traditional Cayley method when the parameter is processing time.
Keywords: matrices; artificial intelligence; time.

[1] Bacharel e Licenciado em Matemática, Bacharel em Ciências Econômicas, Doutor, IFB. fabio.carlucci@ifb.edu.br

Introdução

Este trabalho está focado em um dos papéis da inteligência artificial (IA) na álgebra linear. Mais precisamente, quais as implicações no ajuste de uma função do primeiro grau, aos dados observados, obtida pelo método dos mínimos quadrados ordinários para a regressão linear com abordagem matricial, se o produto de matrizes pode ser realizado por algoritmos obtidos com o auxílio da inteligência artificial.

Na seção 1 da metodologia apresentou-se breve revisão de matrizes e operações de interesse. A seguir, na seção 2, foi apresentado um método de aproximações de funções e o método dos mínimos quadrados (OLS - Ordinary Least Square). A seguir na seção 3, foi apresentado um modelo de regressão linear multivariada, por meio do cálculo matricial (álgebra linear) e, a seguir, uma aplicação. Na Seção 4, foi apresentado o método de Strassen para multiplicação de matrizes, e ainda, foram citados os métodos de multiplicação que utilizam a Inteligência Artificial (IA), em particular o AlphaTensor, da DeepMind, que acarretam diminuição no tempo de processamento para grandes bancos de dados e, consequentemente, maior economia dos fatores de produção. Por fim, na seção 5 apresenta-se um exemplo de aplicação do método de Strassen.

O aprendizado de máquina supervisionado, conhecido também como machine learning supervisionado, é uma subcategoria de machine learning e da inteligência artificial. É definido pelo uso de conjuntos de dados rotulados para treinar algoritmos que classificam dados ou preveem resultados com precisão. O objetivo geral desta pesquisa é proceder a uma investigação a respeito de algoritmos de aprendizado de máquina supervisionados por meio da álgebra matricial.

Referencial teórico

Nas referências deste trabalho, a álgebra linear e álgebra matricial, incluindo a diferenciação matricial, foram consultados: (Boldrini et al.,1980); (Chiang, 1984); (Gujarat, 2006); (Lima,1995), (Lipschutz,1994); (Mingoti, 2007). Por outro lado, para o método dos mínimos quadrados ordinários (OLS – Ordinary Least Square) ou mínimos quadrados multivariados escalares, obtidos por meio do método de aproximação de funções, foram consultados: (Arelanes; Darezzo, 2015); (Boldrini et al.,1980); (Hill.; Griffiths; Judge 2006); (Morettin; Hazzan, Bussab 2009); (Ruggeiro et al., 1996). Para o modelo de regressão linear multivariada comum abordagem da álgebra linear, por meio do cálculo matricial foram consultadas as seguintes referências: (Neter; Wasserman; Kutner,1983), (Gujarati, 2006). Para os conceitos de cálculo, estatística e probabilidade foram consultadas as seguintes referências: (Bussab; Morettin, 2011); (Chung, 1968); (Fávero.; Belfiore 2020); (Leon-Garcia, 1994), (Mingoti, 2007); (Morettin; Hazzan; Bussab, 2009), (Ross, 1983). Para os métodos de Inteligência artificial na álgebra matricial foram consultadas as seguintes referências: Quanta Magazine (2023); MIT Technology Review (2022).

1 Metodologia

Inicialmente será realizada a revisão, multiplicação, inversão e diferenciação de matrizes. A seguir será deduzida uma aproximação de funções pelo método dos mínimos quadrados, tal que:

1. Inicialmente será realizada a revisão, multiplicação, inversão e diferenciação de matrizes.
2. Aproximação de funções: método dos mínimos quadrados.

1.1 Definição.

Uma matriz real $A = (a_{ij})$**,** de ordem $m \times n$ é uma lista de números reais a_{ij} com índices duplos, tais que $1 \leq i \leq m$ e $1 \leq j \leq n$. Essa matriz pode ser representada por uma tabela numérica com m linhas e n colunas, no qual o elemento a_{ij} está localizado na $i-$ é$sima$ linha com a $j-$ é$sima$ coluna:

$$A = [a_{11}\ a_{21}\ \vdots\ a_{n1}\ \ a_{12}\ a_{22}\ \vdots\ a_{n2}\ \ \cdots \cdots \vdots\ \cdots\ \ a_{1n}\ a_{2n}\ \vdots\ a_{mn}\ ]$$

1.2 Multiplicação de matrizes

Sejam A=$[a_{ij}]$ uma matriz de ordem m x n e B=$[b_{ij}]$ uma matriz de ordem n x p. O produto de A por B é uma nova matriz **C**=$[c_{ij}]$ de ordem m x p, tal que, para $i = 1,2, \ldots m$ e $j = 1,2, \ldots, p$, tem-se $c_{ij} = \sum_{k=1}^{n} \quad a_{ik} b_j$. Ou seja, os elementos da matriz C são obtidos multiplicando-se linha por coluna e adicionado esses resultados, em outras palavras cada elemento do produto é o resultado do produto escalar do vetor-linha de A pelo vetor-coluna de B, por exemplo se

$$\mathrm{A} = [2\ 3\ 1\ 5\ 6\ 1\ ] \text{ e } \mathrm{B} = [3\ 1\ 4\ 5\ 5\ 2\ ]\text{, tem-se}$$

$$\mathbf{C}= \mathrm{AB} = [2 \times 3 + 3 \times 4 + 1 \times 5\ 2 \times 1 + 3 \times 5 + 1 \times 2\ 5 \times 3 + 6 \times 4 + 1 \times 5\ 5 \times 1 + 6 \times 5 + 1 \times 2\ ] = [23\ 19\ 44\ 22\ ]$$

1.3 Inversão de matrizes

A inversa de uma matriz quadrada A, de ordem n x n, denotada por A^{-1} , que se lê inversa de A, é uma matriz, caso exista, tal que $AA^{-1} = A^{-1}A = I$**,** onde I é a matriz identidade. de ordem n x n, por exemplo,

$$A = [2\ 4\ 6\ 8\ ] \Rightarrow A^{-1} = [-1{,}00\ 0{,}50\ 0{,}75\ \ -0{,}25\ ]$$

1.4 Diferenciação matricial

Sendo ax o vetor de n variáveis $a_1x_1, a_2x_2, \dots, a_nx_n$, tais que os coeficientes são os números $a_1, a_2, \dots, a_n$, a derivada é o vetor

$$\frac{\partial(ax)}{\partial x} = [a_1\ a_2\ \vdots\ a_n\ ]$$

Por exemplo, sendo x'Ax a forma quadrática tal que

$$x'Ax = [x_1\ \ x_2\ \dots\ \ x_n][a_{11}\ a_{12}\ \dots\ a_{1n}\ \ a_{21}\ a_{22}\ \dots\ a_{2n}\ \ \vdots\ a_{n1}\ \ \vdots\ a_{n2}\ \ \vdots\vdots\ \ \dots\ a_{nn}\ \ ][x_1\ x_2\ \vdots\ x_n\ ]$$

A derivada é

$$\frac{\partial(x'Ax)}{\partial x} = 2Ax$$

que é um vetor coluna de n elementos, ou

$$\frac{\partial(x'Ax)}{\partial x} = 2x'A$$

que é um vetor linha de n elementos.

2. Aproximação de funções

2.1 Generalidades

Nessa seção, todas as funções contínuas reais de variável real, definidas em um intervalo $[a, b]$, com $a \leq b$. O conjunto dessas funções será indicado por $C[a.b]$.

Sejam f, g duas funções pertencentes a $C[a, b]$ e $\omega \in C[a, b]$ uma função fixada, tal que $\omega(x) > 0$, para todo $a < x < b$. A função ω denomina-se função peso. Nessas condições o número real $(f|g)$ definido a seguir

$$(f|g) = \int_a^b \omega(x)f(x)g(x)dx$$

é um produto escalar definido em $C[a, b]$**.**

Sejam f, g duas funções pertencentes a $C[a, b]$ e $\omega_1, \omega_2, \dots, \omega_m$ números reais tais que $\omega_i > 0,$ para $i = 1,2, \dots, m$ e $x_1, x_2, \dots, x_m$ elementos distintos do intervalo $[a, b]$. Nessas condições o definimos o número real $(f|g)$ por

$$(f|g) = \sum_{k=1}^{m} \omega_k f(x_k) g(x_k)$$

onde os números ω_i são os pesos

2.2 O método dos mínimos quadrados (MMQ)

Um caso particular de aproximação de funções é o que trata de ajuste de dados observados a uma função da família G constituída de polinômios de grau igual ou maior do que 1.Por exemplo, desejamos ajustar as despesas com alimentação (y_i) com a renda (x_i) observadas em diversas famílias, de certa localidade, por uma função do primeiro grau (ou uma reta) da forma $y = \beta_1 + \beta_2 x$, onde **y** representa a despesa e **x** representa a renda dessas famílias.

Para estimar β_1 e β_2, é necessário estabelecer uma relação ou fórmula que nos indique como utilizar as observações amostrais. Há muitas possibilidades para resolver essa questão e dentre elas há o princípio dos mínimos quadrados (Ordinary Least Square – OLS).

Por esse princípio devemos procurar uma reta tal que a soma dos quadrados das distâncias verticais dos pontos observados à essa reta seja a menor possível. O intercepto (b_1) e o coeficiente angular (b_2) da reta assim obtida serão as estimativas de β_1 e β_2, respectivamente. Assim, a reta ajustada será $\hat{y} = b_1 + b_2 x$.

De modo que as distâncias dos pontos à reta ou resíduos ou erros aleatórios obtidos são

$$\hat{u}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$$

e a soma dos quadrados dos erros seja a menor possível, ou seja,

$$min \sum \hat{u}_i^{\,2} = min \sum (y_i - \hat{y}_i)^2 = min \sum (y_i - b_1 - b_2 x_i)^2$$

Dessa forma, nosso problema é encontrar o par (b_1, b_2) que minimiza a função

$$f(b_1, b_2) = \sum (y_i - b_1 - b_2 x_i)^2$$

Os pontos críticos dessa função podem ser obtidos estabelecendo-se que

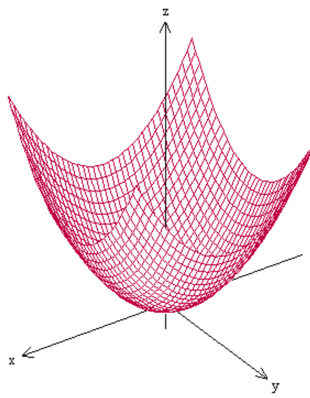
$$\frac{\partial f(b_1, b_2)}{\partial b_1} = \frac{\partial f(b_1, b_2)}{\partial b_2} = 0$$

onde f é uma soma de quadrados de duas variáveis reais, logo f admite somente um ponto de mínimo, não havendo necessidade de se verificar as condições de segunda ordem. Por exemplo, para maiores esclarecimentos, sendo x e y números reais, observemos o gráfico a seguir da função

$$f(x, y) = 0{,}5x^2 + 0{,}5y^2$$

Pelo gráfico, a seguir, pode-se observar que há ponto de mínimo e não há ponto de máximo.

Figura 1 – Ponto de mínimo de $f(x, y) = 0{,}5x^2 + 0{,}5y^2$.



Fonte: Elaboração do Autor.

Em síntese, a reta dos mínimos quadrados tem por equação $\hat{y} = b_1 + b_2x$

onde,

$$b_2 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad \text{e} \quad b_1 = \frac{\sum y_i}{n} - b_2 \frac{\sum x_i}{n}$$

Por exemplo, considere as vendas e os gastos com propaganda de certo produto da Empresa XYZ, cujos valores amostrais estão na Tabela 1 a seguir

Tabela 1 – Vendas e gastos com propaganda da Empresa HTH

Propaganda (x)	30	21	35	42	37	20	8	17	35	25
Vendas (y)	430	335	520	490	470	210	195	270	400	480

Fonte: Elaboração do autor com dados hipotéticos

O coeficiente de correlação é **r = 0,8594**, indicando uma correlação positiva forte. Para obter a equação da reta de regressão, têm-se:

$\sum \quad x_i = 270, \qquad \sum \quad y_i = 3800, \qquad \sum \quad x_i y_i = 112455\,, \qquad \sum \quad x_i^2 = 8302$

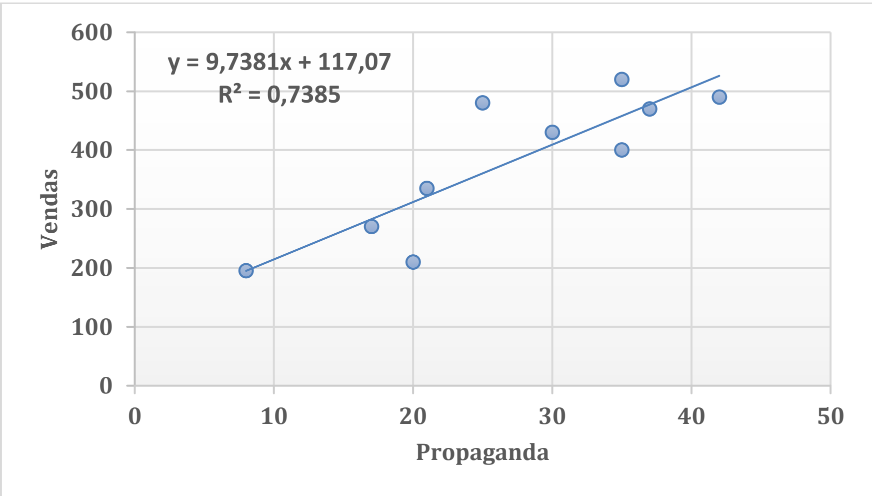
Dessa forma pelas fórmulas, tem-se:

$$b_2 = \frac{112455 - \frac{(270)(3800)}{10}}{8302 - \frac{270^2}{10}} = \frac{9855}{1012} = 9{,}7381 \quad \mathbf{e} \quad b_1 = \frac{3800}{10} - (9{,}7381).\frac{270}{10} = 117{,}0713$$

Portanto, a estimativa da reta procurada é $\hat{y} = 9{,}7381x + 117{,}07$.

Resta verificar se esses resultados satisfazem os pressupostos de uma regressão linear simples, considerando-se a variabilidade amostral. A Figura 2, a seguir, apresenta a reta de regressão linear simples das receitas de vendas em função dos gastos com propaganda.

Figura 2 – Reta de regressão linear simples de vendas e propaganda



Fonte: Elaboração do autor

Tabela 2 – Resultados da regressão

	Coeficientes	*Erro padrão*	*Stat t*	*p-valor*	
Interseção	117,0701581	59,0298508	1,98324	0,08263	
Propaganda (x)	9,738142292	2,048709216	4,75331	0,00144	
ANOVA					
	gl	*SQ*	*MQ*	*F*	*p-valor*
Regressão	1	95969,39229	95969,4	22,5939	0,001439122
Resíduo	8	33980,60771	4247,58		
Total	9	129950			
Soma dos Resíduos	0.00000				

Fonte: Elaboração do autor

Pelo p-valor = 0,00114 do coeficiente da variável explanatória (X = propaganda), verificamos que pode ser aceito seu valor ao nível de significância de 5%. Dessa forma, pelo valor do coeficiente de correlação R, pelo valor do coeficiente de determinação R^2, pela significância do coeficiente da variável exploratória e pela média dos resíduos, podemos aceitar o resultado dessa regressão, de modo que dentro do escopo da variável explicativa, podemos fazer predições. Por exemplo, para um gasto de 40 unidades monetárias em propaganda esse modelo nos indica uma receita de vendas de y = 117,070 + 9,378x40 = 492,19 unidades monetárias.

3 O método de regressão multivariada matricial

Pode-se escrever os modelos de regressão da população com k variáveis, envolvendo a variável target Y e (k-1) variáveis explanatórias $X_2, X_3, \ldots, X_k$ da seguinte forma:

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + u_i \tag{1}$$

onde $i = 1,2,3,\ldots,n$, β_1 é o intercepto e $\beta_2,\ldots,\beta_k$ representam os coeficientes angulares parciais, u_i representam os termos de erros estocásticos, i representa a $i\_ésima$ observação e n representa o tamanho da população.

A interpretação desse modelo é que ele nos informa a média ou valor esperado de Y condicionado aos valores fixos de $X_2, X_{3,}, \ldots, X_k$, obtidos de amostras repetidas, ou seja, a interpretação dessa regressão é $E[Y|X_{2i}, X_{3i}, .\,., X_{ki}]$

A expressão (1) é uma forma abreviada de n equações simultâneas:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 X_{12} + \beta_3 X_{13} + \cdots + \beta_k X_{1k} + u_1 \\ y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{23} + \cdots + \beta_k X_{2k} + u_2 \\ &\text{---} \\ y_n &= \beta_1 + \beta_2 X_{n2} + \beta_3 X_{n3} + \cdots + \beta_k X_{nk} + u_n \end{aligned} \tag{2}$$

Uma alternativa é escrever o modelo (2) na forma matricial, como se segue:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \ldots & x_{1k} \\ 1 & x_{22} & x_{23} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{3}$$

onde y é um vetor coluna de ordem $n \times 1$ de observações da variável Y

X é a matriz $n \times k$ das **n** observações das variáveis $X_2, X_3, \ldots, X_k$

β é o vetor coluna dos coeficientes desconhecidos $\beta_1, \beta_2, \ldots, \beta_k$

u é o vetor coluna $n \times 1$ dos n termos de erro $u_1, u_2, \ldots, u_n$

Dessa forma, conhecendo as ordens das matrizes, podemos reescrever (3) por

$$y = X\beta + u \tag{4}$$

3.1 Pressupostos do modelo de regressão múltipla matricial

Premissa 1. $E[u] = 0$, ou seja, a média dos erros estocásticos deverá ser igual a zero, onde u e 0 são vetores coluna n x 1, sendo 0 o vetor nulo.

Premissa 2. $E[uu'] = \sigma^2 I$, onde I é a matriz identidade.

Premissa 3. A matriz X de ordem n x k não é estocástica, isto é, consiste em um conjunto de números fixos.

Premissa 4. O posto de X é $p(X) = k$, onde k é o número de colunas de X e inferior ao número de linhas.

Premissa 5. $u \sim N(0, \sigma^2 I)$, isto é, o vetor u tem uma distribuição normal multivariada.

Observações:

1. Estamos supondo que não há homocedasticidade e nem correlação serial.
2. A matriz $E[uu'] = \sigma^2 I$ é chamada de matriz de variâncias e covariâncias do termo de erro u.

Supor que a matriz X tem posto igual ao número de colunas é equivalente a supor que as colunas de X são linearmente independentes, em outras palavras não há multicolinearidade.

3.2 Estimativa dos mínimos quadrados ordinários na forma matricial

Para obter uma estimativa para β pelo MQO vamos usar a equação (1 – 4)

$$y = X\hat{\beta} + \hat{u} \Leftrightarrow \hat{u}'.\hat{u} = \left(y - X\hat{\beta}\right)'.\left(y - X\hat{\beta}\right) = \sum \quad \hat{u}^2$$

Como $\left(X\hat{\beta}\right)' = \hat{\beta}'X'$ **e** $\hat{\beta}'X\beta$ é um número escalar, tem-se

$$\hat{u}'.\hat{u} = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

Em notação escalar, diferenciando-se parcialmente em relação a $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ e igualando-se a zero, temos:

$$\frac{\partial\left(\hat{u}'.\hat{u}\right)}{\partial\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 \Leftrightarrow$$

$$(X'X)\hat{\beta} = X'y$$

Aplicando-se a álgebra matricial, se $(X'X)$ é não-singular, tem-se:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Por exemplo, consideremos novamente as vendas e a despesa com propaganda apresentada pela Tabela 1 do exemplo anterior, o modelo de regressão correspondente a esse exemplo hipotético é o modelo de duas variáveis, ou seja, $Y_i = \beta_1 + \beta_2 X_i + u_i$, onde Y

representa a renda e ***X*** representa as despesas semanais com propaganda e ***u*** é um termo de erro. Na forma matricial, temos $y = X\beta + u$ onde

$$_1\, \beta_2 + u_1\, u_2\, u_3\, u_4\, u_5\, u_6\, u_7\, u_8\, u_9\, u_{10}$$

O objetivo é estimar os parâmetros da regressão e obter inferências sobre eles a partir dos dados disponíveis. Nesse exemplo, isto significa calcular $\widehat{\beta}$ e fazer inferência sobre este β. Para realizar o cálculo, podemos usar o método dos produtos e inversões matriciais:
$(X'X) = [10\ 270\ 270\ 8302\ ]$; $(X'y) = [3800\ 112455\ ]$; $(X'X)^{-\prime} = [0{,}820356\ -0{,}026680\ -0{,}026680\ 0{,}000988\ ]$

$$\Rightarrow \hat{\beta} = [117{,}07\ 9{,}7381\ ]$$

Portanto, a equação da reta ajustada de regressão é $\hat{y} = 117{,}07 + 9{,}7381x$. Dessa forma, o modelo de regressão linear clássico, com qualquer número de variáveis explicativas e não estocásticas pode ser resolvido pela abordagem matricial de álgebra linear como uma opção notação escalar que à medida que o número de variáveis aumenta dificulta a notação.

4. A inteligência artificial e o produto de matrizes

4.1 A gênesis do produto matricial e a inteligência artificial

Segundo (Boyer; Merzbach, 2012, p.378), no século XIX, *Arthur Cayley*, Professor de Matemática Pura na Universidade de Cambridge, de 1863 a 1895, estabeleceu o produto de matrizes, como conhecemos hoje. Na época, *Cayley* não estava trabalhando com álgebra, mas sim com geometria. Ele estava interessado em mudança de coordenadas num sistema de coordenadas cartesianas (x, y) para (x’, y’) e de (x’, y’) para (x’’, y’’) e percebeu o seguinte

i) de **(x, y)** para **(x’, y’)**, teria o sistema $\{x' = ax + by\ y' = cx + dy$

ii) de **(x’, y’)** para **(x’’, y’’),** teria o sistema $\{x'' = Ax' + By'\ y'' = Cx' + Dy'$

Portanto, por (i) e (ii), teria os seguintes resultados

$$\{x'' = A(ax + by) + B(cx + dy) = (Aa + Bc)x + (Ab + Bd)y\ y'' = C(ax + by) + D(cx + dy) = (Ca + Dc)x + (Cb + Dd)y$$

Cayley percebeu que esse resultado poderia ser obtido de forma mais prática, usando o conceito de matriz, pois multiplicando-se linha da primeira matriz por coluna da segunda matriz, obtém-se

$$:[A\ B\ C\ D\ ][a\ b\ c\ d\ ][x\ y\ ] = [Aa + Bc\ Ab + Bd\ Ca + Dc\ Cb + Dd\ ][x\ y\ ]$$

Logo, este procedimento foi adotado como o produto usual de duas matrizes, desde que o número de colunas da primeira matriz fosse igual ao número de linhas da segunda matriz. Note que são utilizadas oito multiplicações e quatro adições.

4.2 O uso da inteligência artificial no produto de matrizes

Os métodos numéricos de aparência e formulação teórica aparentemente simples, na realidade podem se tornar muito trabalhosos quando executados manualmente ou até por meio de um computador. Embora o computador seja utilizado em razão de sua velocidade e capacidade de executar automaticamente grande número de instruções, por meio de algoritmos e estruturas de dados, ele é um instrumento que depende da capacidade e inteligência de quem o usa.

Considerando apenas o tempo de execução como um dos fatores entre os que requerem atenção na avaliação de um método, vejamos como esse fator pode variar criticamente para um mesmo problema, de acordo com o método adotado. A multiplicação de matrizes é um cálculo que possui fundamentais aplicações como, por exemplo, exibir imagens em uma tela, simular problemas complexos de física e o próprio machine learning, como o exemplo resolvido do exemplo 3.2. O aumento na velocidade desse cálculo pode causar impacto em milhares de tarefas computacionais do dia a dia, reduzindo gastos e economizando energia.

Segundo a *Quanta Magazine*, de 23 de novembro de 2002, em 1969 o matemático alemão Volker Strassen descobriu um procedimento para a multiplicação de matrizes diferente do método desenvolvido por Cayley, mas com os mesmos resultados. Seu algoritmo é muito simples para uma matriz 2x2 e utiliza somente sete multiplicações.

Logo, Strassen percebeu que o mesmo algoritmo funcionaria para matrizes de ordem bem maior como, por exemplo, uma matriz com 20.000 (vinte mil) linhas e 20.000 (vinte mil colunas), pois essa matriz pode ser obtida de uma matriz 2x2, cujos quatro elementos são matrizes de 10.000 linhas e 10.000 colunas que, por sua vez, cada uma dessas matrizes pode ser subdividida em blocos de 5.000 linhas e 5000 colunas, e assim por diante. Seu algoritmo poderia aplicar seu método de multiplicar matrizes 2x2 em cada nível dessa hierarquia aninhada.

Por outro lado, segundo a *MIT Technology Review*, de outubro de 2022, a Inteligência Artificial (IA) com a *AlphaZero*, treinada para jogos de tabuleiro, foi usada pela empresa

DeepMind, para descobrir maneiras mais rápidas de multiplicar matrizes, batendo um recorde de 50 anos na ciência da computação.

Segundo Thomas Hubert, engenheiro da *DeepMind*, existem mais maneiras de multiplicar duas matrizes do que 10 elevado à potência 33. O truque foi transformar o problema em um jogo de tabuleiro tridimensional chamado TensorGame. Os pesquisadores da DeepMind treinaram uma versão da AlphaZero, chamada *AlphaTensor*, para jogar esse jogo (multiplicar matrizes). A multiplicação de duas matrizes 4 x 4 pelo método tradicional tem 64 passos, o mesmo resultado, pelo método de Strassem, tem 49 passos e o programa AlphaTensor descobriu uma maneira que obtém o mesmo resultado em 47 passos.

4.3 Apresenta-se a seguir o algoritmo de Strassen para o produto de matrizes 2 x 2

Considere as matrizes 2x2 a seguir:

X = [1 2 3 4] e Y = [5 6 7 8]

Pelo método de Cayley o produto de A por B é

XY = [1 2 3 4][5 6 7 8] = [$1x5 + 2x7$ $1x6 + 2x8$ $3x5 + 4x7$ $3x6 + 4x8$] = [19 22 43 50]

São oito multiplicações e 4 adições. Por outro lado, pelo método de Strassen, temos os seguintes cálculos:

$$(x_{11} + x_{22}) \times (y_{11} + y_{22}) = (1 + 4) \times (5 + 8) = 65 = A$$
$$(x_{21} + x_{22}) \times y_{11} = (3 + 4) \times 5 = 35 = B$$
$$(y_{12} - y_{22}) \times x_{11} = (6 - 8) \times 1 = -2 = C$$
$$(-y_{11} + y_{21}) \times x_{22} = (-5 + 7) \times 4 = 8 = D$$
$$(x_{11} + x_{12}) \times y_{22} = (1 + 2) \times 8 = 24 = E$$
$$(-x_{11} + x_{21}) \times (y_{11} + y_{12}) = (-1 + 3) \times (5 + 6) = 22 = F$$
$$(x_{12} - x_{22}) \times (y_{21} + y_{22}) = (2 - 4) \times (7 + 8) = -30 = G$$

Solução:

$$x_{11} = A + D - E + G = 65 + 8 - 24 - 30 = 19$$
$$x_{12} = C + E = -2 + 24 = 22$$
$$x_{21} = B + D = 35 + 8 = 43$$
$$x_{22} = A - B + C + F = 65 - 35 + (-2) + 22 = 50$$

São sete multiplicações e 10 adições.

XY = [1 2 3 4][5 6 7 8] = [19 22 43 50]

4.4 Comparação do produto matricial tradicional com o do método de Strassen

O autor, usando o mesmo Laptop da HP, Intel Corel I3, com Windows 10, realizou o produto de matrizes pelo método tradicional de Cayley e, a seguir, pelo novo método de Strassen, para observar o tempo de processamento em cada caso.

4.4.1 Produto de matrizes pelo método de Cayley

X =[19 22 43 50], elapsed time is 3.90655172 seconds.

4.4.2 Produto de matrizes realizado com o método de Strassen

X =[19 22 43 50], elapsed time is 0.000284 seconds.

5. Conclusão

Neste trabalho, realizou-se um estudo dos modelos de regressão linear clássico envolvendo k+1 variáveis (Y e X_1, X_2, X_3, ..., X_k) por meio da álgebra linear. Conceitualmente, o modelo de k +1 variáveis é uma extensão dos modelos de duas e três variáveis por meio do método dos mínimos quadrados (MQO). A vantagem da abordagem por álgebra linear sobre o método escalar é utilizar o método matricial, por ser um método compacto para lidar com modelos de regressão linear envolvendo grande número de variáveis. Na introdução, foram estabelecidos os objetivos desse trabalho, assim como os materiais e métodos para seu desenvolvimento. Na seção 1, foi realizada revisão breve de matrizes e operações, como estudadas no ensino médio, exceto a diferenciação que exige conhecimento de cálculo diferencial.

Na seção 2, foi desenvolvido o modelo de regressão linear simples, por meio do método dos mínimos quadrados ordinários escalar. A seguir, na seção 3, foi introduzido o modelo de regressão linear multivariada com abordagem da álgebra linear, por meio do cálculo matricial. Foi realizado um exemplo ilustrativo desse método. Por fim, na seção 4, foi realizado um breve histórico sobre uma nova forma para a multiplicação de matrizes, utilizando o método de Strassen e foram citados os métodos de multiplicação que utilizam a Inteligência Artificial (IA), em particular o AlphaTensor, da DeepMind, que acarretam diminuição no tempo de processamento para grandes bancos de dados e, consequentemente, maior economia dos fatores de produção.

Referências

ARENALES, Selma; DAREZZO, Artur. **Cálculo Numérico**: aprendizagem com apoio de software. São Paulo: Cengage Learning, 2015.

BARROS, Ivan de Queiroz. **Introdução ao cálculo numérico**, São Paulo: Editora Edgard Blücher; Editora da USP, 1972.

BOUDRINI, José Luiz et al., **Álgebra Linear**, São Paulo: Harper & Row do Brasil, 1980.

BOYER, Carl; MERZBACH, UTA C. **História da Matemática**. São Paulo: Editora Blücher, 2012.

CHUNG, K.L. **A Course in Probability Theory**. New York: Harcourt, Brace & World, Inc., 1968.

FÁVERO, L. P.; BELFIORE, P**. Manual de Análise de Dados**. Rio de Janeiro: LTC, 2020.

GUJARATI, D. **Econometria Básica**. Rio de Janeiro: Editora Campus, 2006.

HADLEY, G. **Linear Algebra**. Reading, Mass: Addison-Wesley, 1961.

HILL, R.C; GRIFFITHS, W.E.; JUDGE, G.G. 2 ed. **Econometria.** São Paulo: Editora Saraiva, 2006.

MIT Technology Review. **A Inteligência Artificial de DeepMind bateu um Record de 50 anos na ciência da Computação,** 28 de outubro de 2022.Disponível em: <https://deepmind.google/discover/blog/discovering-**novel**-algorithms-with-alphatensor/.>. Acesso em: nov.2023.

NETER, John; WASSERMAN, William; KUTNER, Michael H. **Applied Linear Regression Models**. New York: Richard Irwin, Inc. 1983.

STRASSEN, Volker. Algorithm. **Quanta Magazine**. Disponível em:<https//www.quantamagazine.org/ai-reveals-new-possibilities-in-matrix- multiplication-2022-1123/>. Acessado em nov.2023.