www.datascienceacademy.com.br

Matemática Para Machine Learning

Operações com Matrizes
Multiplicação de Matrizes

Vimos as operações element-wise e na verdade esse tipo de operação cobre muito do que é feito no pré-processamento de dados para diversos modelos de Machine Learning. Mas quando precisamos multiplicar matrizes, as coisas ficam um pouco mais complicadas. Vejamos como isso funciona.

Aqui temos duas matrizes e queremos realizar a multiplicação entre elas.

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \times \begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} = \begin{bmatrix}1 & 4 & 9\\16 & 25 & 36\\49 & 64 & 81\end{bmatrix}$$

O resultado deixa claro como a operação de multiplicação foi realizada e na prática realizamos uma operação element-wise, totalmente válida.

$$\begin{bmatrix}\boxed{1} & \boxed{2} & \boxed{3}\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \times \begin{bmatrix}\boxed{1} & \boxed{2} & \boxed{3}\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} = \begin{bmatrix}\boxed{1} & \boxed{4} & \boxed{9}\\16 & 25 & 36\\49 & 64 & 81\end{bmatrix}$$

Entretanto, quando realizamos operações de multiplicação entre matrizes, normalmente outro método é usado, chamado “matrix products”, gerando um resultado como demonstrado na figura abaixo:

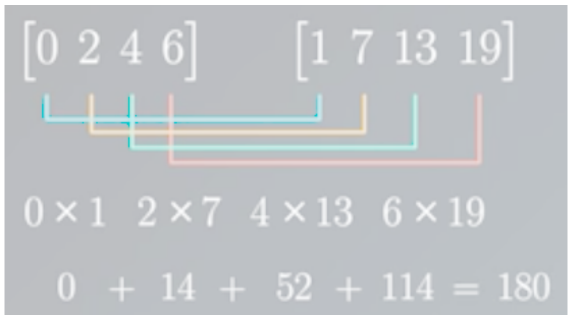
$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \cdot \begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} = \begin{bmatrix}30 & 36 & 42\\66 & 81 & 96\\102 & 126 & 150\end{bmatrix}$$

Mas como chegamos a esse resultado? Continue acompanhando.

Antes de voltar as matrizes, vamos realizar a operação de multiplicação entre 2 vetores com o mesmo shape (mesmas dimensões):

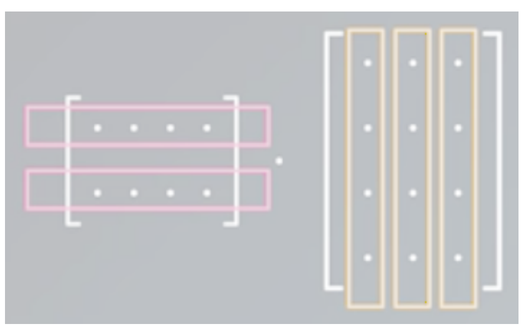
$$\begin{bmatrix} 0 & 2 & 4 & 6 \end{bmatrix} \quad \begin{bmatrix} 1 & 7 & 13 & 19 \end{bmatrix}$$

Uma operação que podemos realizar com estes vetores, é chamado de "dot product". Para encontrar o "dot product", primeiro multiplicamos os elementos correspondentes em cada vetor e depois somamos os resultados, conforme imagem abaixo (você estudou isso no capítulo anterior):



Ou seja, multiplicando 2 vetores de qualquer comprimento, mas que tenham o mesmo shape, nos dá como resultado um valor escalar, nesse caso um único número, 180.

E como encontramos então o "dot product" de duas matrizes? Simples. Aplicamos o mesmo conceito que usamos nos vetores, encontrando o "dot product" entre uma linha da primeira matriz, com uma coluna da segunda matriz (na prática é como se estivéssemos encontrando o dot product entre 2 vetores).



Quando estamos multiplicando 2 matrizes, estamos operando as linhas de uma matriz com as colunas da outra matriz:

$$\begin{bmatrix} 0 & 2 & 4 & 6 \\ 8 & 10 & 12 & 14 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \\ 19 & 21 & 23 \end{bmatrix}$$

Com isso, tudo que precisamos fazer é encontrar o dot product entre a linha da primeira matriz e a coluna correspondente na segunda matriz e assim por diante.

$$\begin{bmatrix} 0 & 2 & 4 & 6 \\ 8 & 10 & 12 & 14 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \\ 19 & 21 & 23 \end{bmatrix} = \begin{bmatrix} 180 & & \\ & & \end{bmatrix}$$

Perceba que a matriz resultante terá o mesmo número de linhas da primeira matriz e o mesmo número de colunas da segunda matriz. Para encontrar a matriz resultante, foram necessárias 24 multiplicações e 18 adições, tudo isso para encontrar os 6 números finais. Imagine grandes conjuntos de dados. A multiplicação de matrizes vai requerer muito poder computacional e muitas tarefas repetitivas serão realizadas. Por isso as GPUs são indicadas para este tipo de operação. **Quando trabalhamos com reconhecimento de imagens, estamos essencialmente trabalhando com matrizes.**

$$\begin{bmatrix} 0 & 2 & 4 & 6 \\ 8 & 10 & 12 & 14 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \\ 19 & 21 & 23 \end{bmatrix} = \begin{bmatrix} 180 & 204 & 228 \\ 500 & 588 & 676 \end{bmatrix}$$

Lembretes importantes sobre a multiplicação de Matrizes:

- O número de colunas na matriz esquerda deve ser igual ao número de linhas na matriz direita.
- A matriz de respostas sempre possui o mesmo número de linhas que a matriz esquerda e o mesmo número de colunas que a matriz direita.
- Ordem importa aqui. Multiplicar A • B não é o mesmo que multiplicar B • A.
- Os dados na matriz esquerda devem ser organizados como linhas, enquanto os dados na matriz direita devem ser organizados como colunas.

Se você tiver estes quatro pontos em mente, você sempre irá descobrir como organizar adequadamente as suas multiplicações de matriz ao construir um modelo de Machine Learning. No Jupyter Notebook anexo veremos como fazer isso usando o NumPy.

Mas e quando as matrizes não estiverem no shape ideal? Para isso, podemos usar a transposta da matriz. Mas fique tranquilo, chegaremos lá ainda neste capítulo. Agora, pratique um pouco a multiplicação de matrizes.