## Os grandes volumes de dados disponíveis para as organizações atualmente tornaram possíveis diversos recursos de inteligência artificial (IA) que antes pareciam ficção científica. Na área de TI, falamos durante anos sobre “big data” e os desafios que as empresas enfrentam para descobrir como processar e utilizar todos os seus dados. A maioria deles, cerca de 80%, não são estruturados, portanto, algoritmos tradicionais não são capazes de analisá-los.knlnkkkmnnin j j j omlmimilmk,mklmk, ,i ik k j

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## kjknmjlnklknlknnoklnkn nn,,m, ,m , m,,

## Algumas década

## s atrás, os pesquisadores descobriram as redes neurais, algoritmos*deep learning*, que podem revelar insights de dados e, muitas vezes, são insights que nunca poderíamos imaginar. Se pudermos executar esses algoritmos dentro de um cronograma viável, eles poderão ser utilizados para analisar nossos dados e revelar padrões dentro deles, o que, por sua vez, pode ajudar nas decisões de negócios. Entretanto, esses algoritmos exigem muita capacidade de computação.

## Treinando redes neurais

## O algoritmo de *deep learning* é aquele que utiliza rede neural para resolver um determinado problema. Uma rede neural é um tipo de algoritmo de IA que usa uma entrada, faz ela atravessar sua rede de neurônios, chamada de camadas, e apresenta uma saída. Quanto mais camadas de neurônios ela tiver, mais profunda será a rede. Se a saída estiver correta, ótimo. Se a saída estiver incorreta, é esperado que o algoritmo verá e “adaptará” as conexões dos neurônios de maneira que, da próxima vez que você fornecer essa entrada, ela apresentará a resposta correta.

## 

## 

## *Figura 1: Ilustração das redes neurais de comput*

## 

## Essa capacidade de treinar uma rede neural até que ela aprenda a apresentar a resposta correta é um aspecto importante da computação cognitiva. As redes neurais aprendem com os dados aos quais elas são expostas e reorganizam a conexão entre os neurônios.

## As conexões entre os neurônios são outro aspecto importante, e a intensidade da conexão entre eles pode variar (ou seja, a ligação pode ser forte, fraca ou ter uma intensidade intermediária). Portanto, quando uma rede neural se adapta, ela está de fato ajustando a intensidade das conexões entre seus neurônios e assim, da próxima vez, ela poderá fornecer uma resposta mais precisa. Para que uma rede neural forneça uma boa resposta para um problema, essas conexões precisam ser ajustadas fazendo o treinamento exaustivo e repetido da rede, ou seja, expondo-a a dados. Pode haver zilhões de neurônios envolvidos, e o ajuste de suas conexões é um procedimento matemático baseado em matriz que faz uso intensivo de computação.

## Precisamos da força dos dados e da computação

## De acordo com discussões atuais, a maioria das organizações tem grandes volumes de dados que podem ser utilizados para treinar essas redes neurais. Porém, ainda existe o problema de toda a matemática massiva e intensiva necessária para calcular as conexões de neurônios durante o treinamento. Tão poderosas quanto os processadores atuais, elas podem executar inúmeras operações matemáticas por segundo. Uma rede neural com zilhões de neurônios treinados por milhares de iterações de treinamento ainda exigirá um zilhão de operações para ser calculada. E agora?

## Graças aos avanços no setor (e eu acredito que a indústria de jogos desempenhou um papel fundamental aqui), há algumas peças de hardware que são excelentes para lidar com operações baseadas em matriz chamadas Graphics Processing Unit (GPU). O GPU pode calcular zilhões de pixels em operações do tipo matriz para mostrar gráficos de alta qualidade em uma tela. E, ao que parece, o GPU pode trabalhar em operações matemáticas da rede neural da mesma maneira.

## Deixe-me apresentar nosso melhor aluno de matemática da sala: o GPU!