



**Data Science  
Academy**

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

Machine Learning

Deep Learning



O interesse pela Aprendizagem de Máquina (Machine Learning) explodiu na última década. O mundo a nossa volta está passando por uma transformação e vemos uma interação cada vez maior das aplicações de computador com os seres humanos. Softwares de detecção de spam, sistemas de recomendação, marcação em fotos de redes sociais, assistentes pessoais ativados por voz, carros autônomos, smartphones com reconhecimento facial e muito mais.

E o interesse por Machine Learning se mostra ainda mais evidente pelo número cada vez maior de conferências, meetups, artigos, livros, cursos, buscas no Google e profissionais e empresas procurando compreender o que é e como usar aprendizagem de máquina, embora muitos ainda confundem o que podem fazer com o que desejam fazer. Não há como ficar indiferente a esta revolução trazida pela aprendizagem de máquina e, segundo o Gartner, até 2020 todos os softwares corporativos terão alguma funcionalidade ligada a Machine Learning.

Fundamentalmente, Machine Learning é a utilização de algoritmos para extrair informações de dados brutos e representá-los através de algum tipo de modelo matemático. Usamos então este modelo para fazer inferências a partir de outros conjuntos de dados. Existem muitos algoritmos que permitem fazer isso, mas um tipo em especial vem se destacando, as redes neurais artificiais.

As redes neurais artificiais não são necessariamente novas, existem pelo menos desde a década de 1950. Mas durante várias décadas, embora a arquitetura desses modelos tivesse evoluído, ainda faltavam ingredientes que fizessem os modelos realmente funcionar. E esses ingredientes surgiram quase ao mesmo tempo. Um deles você já deve ter ouvido: Big Data. O volume de dados, gerado em variedade e velocidade cada vez maiores, permite criar modelos e atingir altos níveis de precisão. Mas ainda falta um ingrediente. Faltava! Como processar grandes modelos de Machine Learning com grandes quantidades de dados? As CPUs não conseguiam dar conta do recado.

Foi quando os gamers e sua afeição por poder computacional e gráficos perfeitos, nos ajudaram a encontrar o segundo ingrediente: Programação Paralela em GPUs. As unidades de processamento gráfico, que permitem realizar operações matemáticas de forma paralela, principalmente operações com matrizes e vetores, elementos presentes em modelos de redes neurais artificiais, formaram a tempestade perfeita, que permitiu a evolução na qual nos encontramos hoje: Big Data + Processamento Paralelo + Modelos de Aprendizagem de Máquina = Inteligência Artificial.

A unidade fundamental de uma rede neural artificial é um nó (ou neurônio matemático), que por sua vez é baseado no neurônio biológico. As conexões entre esses neurônios matemáticos também foram inspiradas em cérebros biológicos, especialmente na forma como essas conexões se desenvolvem ao longo do tempo com “treinamento”. Em meados da década de 1980 e início da década de 1990, muitos avanços importantes na arquitetura das redes neurais artificiais ocorreram. No entanto, a quantidade de tempo e dados necessários para obter bons resultados retardou a adoção e, portanto, o interesse foi arrefecido, com o que ficou conhecido como AI Winter (Inverno da IA).



No início dos anos 2000, o poder computacional expandiu exponencialmente e o mercado viu uma “explosão” de técnicas computacionais que não eram possíveis antes disso. Foi quando o aprendizado profundo (Deep Learning) emergiu do crescimento computacional explosivo dessa década como o principal mecanismo de construção de sistemas de Inteligência Artificial, ganhando muitas competições importantes de aprendizagem de máquina. O interesse por Deep Learning não para de crescer e hoje vemos o termo aprendizado profundo sendo mencionado com frequência cada vez maior e soluções comerciais surgindo a todo momento.

Aprendizagem Profunda ou Deep Learning, é uma subárea da Aprendizagem de Máquina, que emprega algoritmos para processar dados e imitar o processamento feito pelo cérebro humano. Não se preocupe se alguns termos mais técnicos não fizerem sentido agora. Todos eles serão estudados no capítulo sobre Deep Learning.

Deep Learning usa camadas de neurônios matemáticos para processar dados, compreender a fala humana e reconhecer objetos visualmente. A informação é passada através de cada camada, com a saída da camada anterior fornecendo entrada para a próxima camada. A primeira camada em uma rede é chamada de camada de entrada, enquanto a última é chamada de camada de saída. Todas as camadas entre as duas são referidas como camadas ocultas. Cada camada é tipicamente um algoritmo simples e uniforme contendo um tipo de função de ativação.

A aprendizagem profunda é responsável por avanços recentes em visão computacional, reconhecimento de fala, processamento de linguagem natural e reconhecimento de áudio. O aprendizado profundo é baseado no conceito de redes neurais artificiais, ou sistemas computacionais que imitam a maneira como o cérebro humano funciona.

A extração de recursos é outro aspecto da Aprendizagem Profunda. A extração de recursos usa um algoritmo para construir automaticamente “recursos” significativos dos dados para fins de treinamento, aprendizado e compreensão. Normalmente, o Cientista de Dados, ou Engenheiro de IA, é responsável pela extração de recursos.

O aumento rápido e o aparente domínio do aprendizado profundo sobre os métodos tradicionais de aprendizagem de máquina em uma variedade de tarefas tem sido surpreendentes de testemunhar e, às vezes, difícil de explicar. Deep Learning é uma evolução das Redes Neurais, que por sua vez possuem uma história fascinante que remonta à década de 1940, cheia de altos e baixos, voltas e reviravoltas, amigos e rivais, sucessos e fracassos. Em uma história digna de um filme dos anos 90, uma ideia que já foi uma espécie de patinho feio floresceu para se tornar a bola da vez.

Consequentemente, o interesse em aprendizagem profunda tem disparado, com cobertura constante na mídia popular. A pesquisa de aprendizagem profunda agora aparece rotineiramente em revistas como Science, Nature, Nature Methods e Forbes apenas para citar alguns. O aprendizado profundo conquistou Go, aprendeu a dirigir um carro, diagnosticou



câncer de pele e autismo, tornou-se um falsificador de arte mestre e pode até alucinar imagens fotorrealistas.

Os primeiros algoritmos de aprendizagem profunda que possuíam múltiplas camadas de características não-lineares podem ser rastreados até Alexey Grigoryevich Ivakhnenko (desenvolveu o Método do Grupo de Manipulação de Dados) e Valentin Grigor'evich Lapa (autor de Cybernetics and Forecasting Techniques) em 1965, que usaram modelos finos mas profundos com funções de ativação polinomial os quais eles analisaram com métodos estatísticos. Em cada camada, eles selecionavam os melhores recursos através de métodos estatísticos e encaminhavam para a próxima camada. Eles não usaram Backpropagation para treinar a rede de ponta a ponta, mas utilizaram mínimos quadrados camada-por-camada, onde as camadas anteriores foram independentemente instaladas em camadas posteriores (um processo lento e manual).

No final da década de 1970, o primeiro inverno da IA começou, resultado de promessas que não poderiam ser mantidas. O impacto desta falta de financiamento limitou a pesquisa em Redes Neurais Profundas e Inteligência Artificial. Felizmente, houve indivíduos que realizaram a pesquisa sem financiamento.

As primeiras “redes neurais convolutivas” foram usadas por Kunihiro Fukushima. Fukushima concebeu redes neurais com múltiplas camadas de agrupamento e convoluções. Em 1979, ele desenvolveu uma rede neural artificial, chamada Neocognitron, que usava um design hierárquico e multicamadas. Este design permitiu ao computador “aprender” a reconhecer padrões visuais. As redes se assemelhavam a versões modernas, mas foram treinadas com uma estratégia de reforço de ativação recorrente em múltiplas camadas, que ganhou força ao longo do tempo. Além disso, o design de Fukushima permitiu que os recursos importantes fossem ajustados manualmente aumentando o “peso” de certas conexões.

Muitos dos conceitos de Neocognitron continuam a ser utilizados. O uso de conexões de cima para baixo e novos métodos de aprendizagem permitiram a realização de uma variedade de redes neurais. Quando mais de um padrão é apresentado ao mesmo tempo, o Modelo de Atenção Seletiva pode separar e reconhecer padrões individuais deslocando sua atenção de um para o outro (o mesmo processo que usamos em multitarefa). Um Neocognitron moderno não só pode identificar padrões com informações faltantes (por exemplo, um número 5 desenhado de maneira incompleta), mas também pode completar a imagem adicionando as informações que faltam. Isso pode ser descrito como “inferência”.

O Backpropagation, o uso de erros no treinamento de modelos de Deep Learning, evoluiu significativamente em 1970. Foi quando Seppo Linnainmaa escreveu sua tese de mestrado, incluindo um código FORTRAN para Backpropagation. Infelizmente, o conceito não foi aplicado às redes neurais até 1985. Foi quando Rumelhart, Williams e Hinton demonstraram o Backpropagation em uma rede neural que poderia fornecer representações de distribuição “interessantes”. Filosoficamente, essa descoberta trouxe à luz a questão dentro da psicologia cognitiva de saber se a compreensão humana depende da lógica simbólica

(computacionalismo) ou de representações distribuídas (conexão). Em 1989, Yann LeCun forneceu a primeira demonstração prática de Backpropagation no Bell Labs. Ele combinou redes neurais convolutivas com Backpropagation para ler os dígitos “manuscritos” (assunto do próximo capítulo). Este sistema foi usado para ler o número de cheques manuscritos.



Os padrinhos da Inteligência Artificial. Da esquerda para a direita: Yann LeCun, Geoffrey Hinton, Yoshua Bengio e Andrew Ng

Porém, tivemos neste período o que ficou conhecido como segundo Inverno da IA, que ocorreu entre 1985-1990, que também afetou pesquisas em Redes Neurais e Aprendizagem Profunda. Vários indivíduos excessivamente otimistas haviam exagerado o potencial “imediato” da Inteligência Artificial, quebrando as expectativas e irritando os investidores. A raiva era tão intensa, que a frase Inteligência Artificial atingiu o status de pseudociência. Felizmente, algumas pessoas continuaram trabalhando em IA e Deep Learning, e alguns avanços significativos foram feitos. Em 1995, Dana Cortes e Vladimir Vapnik desenvolveram a máquina de vetor de suporte ou Support Vector Machine (um sistema para mapear e reconhecer dados semelhantes). O LSTM (Long-Short Term Memory) para redes neurais recorrentes foi desenvolvido em 1997, por Sepp Hochreiter e Juergen Schmidhuber.



O próximo passo evolutivo significativo para Deep Learning ocorreu em 1999, quando os computadores começaram a se tornar mais rápidos no processamento de dados e GPUs (unidades de processamento de gráfico) foram desenvolvidas. O uso de GPUs significou um salto no tempo de processamento, resultando em um aumento das velocidades computacionais em 1000 vezes ao longo de um período de 10 anos. Durante esse período, as redes neurais começaram a competir com máquinas de vetor de suporte. Enquanto uma rede neural poderia ser lenta em comparação com uma máquina de vetor de suporte, as redes neurais ofereciam melhores resultados usando os mesmos dados. As redes neurais também têm a vantagem de continuar a melhorar à medida que mais dados de treinamento são adicionados.

Em torno do ano 2000, apareceu o problema conhecido como Vanishing Gradient. Foi descoberto que as “características” aprendidas em camadas mais baixas não eram aprendidas pelas camadas superiores, pois nenhum sinal de aprendizado alcançou essas camadas. Este não era um problema fundamental para todas as redes neurais, apenas aquelas com métodos de aprendizagem baseados em gradientes. A origem do problema acabou por ser certas funções de ativação. Uma série de funções de ativação condensavam sua entrada, reduzindo, por sua vez, a faixa de saída de forma um tanto caótica. Isso produziu grandes áreas de entrada mapeadas em uma faixa extremamente pequena. Nessas áreas de entrada, uma grande mudança será reduzida a uma pequena mudança na saída, resultando em um gradiente em queda. Duas soluções utilizadas para resolver este problema foram o pré-treino camada-a-camada e o desenvolvimento de uma memória longa e de curto prazo.

Em 2001, um relatório de pesquisa do Grupo META (agora chamado Gartner) descreveu os desafios e oportunidades no crescimento do volume de dados. O relatório descreveu o aumento do volume de dados e a crescente velocidade de dados como o aumento da gama de fontes e tipos de dados. Este foi um apelo para se preparar para a investida do Big Data, que estava apenas começando.

Em 2009, Fei-Fei Li, professora de IA em Stanford na Califórnia, lançou o ImageNet e montou uma base de dados gratuita de mais de 14 milhões de imagens etiquetadas. Eram necessárias imagens marcadas para “treinar” as redes neurais. A professora Li disse: “Nossa visão é que o Big Data mudará a maneira como a aprendizagem de máquina funciona. Data drives learning.”. Ela acertou em cheio!

Até 2011, a velocidade das GPUs aumentou significativamente, possibilitando a formação de redes neurais convolutivas “sem” o pré-treino camada por camada. Com o aumento da velocidade de computação, tornou-se óbvio que Deep Learning tinha vantagens significativas em termos de eficiência e velocidade. Um exemplo é a AlexNet, uma rede neural convolutiva, cuja arquitetura ganhou várias competições internacionais durante 2011 e 2012. As unidades lineares retificadas foram usadas para melhorar a velocidade.

Também em 2012, o Google Brain lançou os resultados de um projeto incomum conhecido como The Cat Experiment. O projeto de espírito livre explorou as dificuldades de





“aprendizagem sem supervisão”. A Aprendizagem profunda usa “aprendizagem supervisionada”, o que significa que a rede neural convolutiva é treinada usando dados rotulados. Usando a aprendizagem sem supervisão, uma rede neural convolucional é alimentada com dados não marcados, e é então solicitada a busca de padrões recorrentes.

O Cat Experiment usou uma rede neural distribuída por mais de 1.000 computadores. Dez milhões de imagens “sem etiqueta” foram tiradas aleatoriamente do YouTube, mostradas ao sistema e, em seguida, o software de treinamento foi autorizado a ser executado. No final do treinamento, um neurônio na camada mais alta foi encontrado para responder fortemente às imagens de gatos. Andrew Ng, o fundador do projeto, disse: “Nós também encontramos um neurônio que respondeu fortemente aos rostos humanos”. A aprendizagem não supervisionada continua a ser um campo ativo de pesquisa em Aprendizagem Profunda.

Atualmente, o processamento de Big Data e a evolução da Inteligência Artificial são ambos dependentes da Aprendizagem Profunda. Com Deep Learning podemos construir sistemas inteligentes e estamos nos aproximando da criação de uma IA totalmente autônoma. Isso vai gerar impacto em todas as partes da sociedade e aqueles que souberem trabalhar com a tecnologia, serão os líderes desse novo mundo que se apresenta diante de nós.

Caso queira se aprofundar no estudo de Deep Learning, acompanhe um projeto fascinante mantido pela DSA gratuitamente, o Deep Learning Book:

<http://deeplearningbook.com.br/>