

Deep Learning Book

Data Science Academy

Capítulo 1 – Deep Learning e a Tempestade Perfeita

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.

Capítulo 2 – Uma Breve História das Redes Neurais Artificiais

O desenvolvimento do cérebro humano ocorre principalmente nos dois primeiros anos de vida, mas se arrasta por toda a vida. Inspirando-se neste modelo, diversos pesquisadores tentaram simular o funcionamento do cérebro, principalmente o processo de aprendizagem por experiência, a fim de criar sistemas inteligentes capazes de realizar tarefas como classificação, reconhecimento de padrões, processamento de imagens, entre outras

atividades. Como resultado destas pesquisas surgiu o modelo do neurônio artificial e posteriormente um sistema com vários neurônios interconectados, a chamada Rede Neural.

Em 1943, o neurofisiologista Warren McCulloch e o matemático Walter Pitts escreveram um artigo sobre como os neurônios poderiam funcionar e para isso, eles modelaram uma rede neural simples usando circuitos elétricos.

Warren McCulloch e Walter Pitts criaram um modelo computacional para redes neurais baseadas em matemática e algoritmos denominados lógica de limiar (threshold logic). Este modelo abriu o caminho para a pesquisa da rede neural dividida em duas abordagens: uma abordagem focada em processos biológicos no cérebro, enquanto a outra focada na aplicação de redes neurais à inteligência artificial.

Em 1949, Donald Hebb escreveu *The Organization of Behavior*, uma obra que apontou o fato de que os caminhos neurais são fortalecidos cada vez que são usados, um conceito fundamentalmente essencial para a maneira como os humanos aprendem. Se dois nervos dispararem ao mesmo tempo, argumentou, a conexão entre eles é melhorada.

À medida que os computadores se tornaram mais avançados na década de 1950, finalmente foi possível simular uma hipotética rede neural. O primeiro passo para isso foi feito por Nathaniel Rochester dos laboratórios de pesquisa da IBM. Infelizmente para ele, a primeira tentativa de fazê-lo falhou.

No entanto, ao longo deste tempo, os defensores das “máquinas pensantes” continuaram a argumentar suas pesquisas. Em 1956, o Projeto de Pesquisa de Verão de Dartmouth sobre Inteligência Artificial proporcionou um impulso tanto à Inteligência Artificial como às Redes Neurais. Um dos resultados deste processo foi estimular a pesquisa em IA na parte de processamento neural.

Nos anos seguintes ao Projeto Dartmouth, John von Neumann sugeriu imitar funções simples de neurônios usando relés telegráficos ou tubos de vácuo. Além disso, Frank Rosenblatt, um neurobiologista, começou a trabalhar no Perceptron. Ele estava intrigado com o funcionamento do olho de uma mosca. Grande parte do processamento feito por uma mosca ao decidir fugir, é feito em seus olhos. O Perceptron, que resultou dessa pesquisa, foi construído em hardware e é a mais antiga rede neural ainda em uso hoje. Um Perceptron de camada única foi útil para classificar um conjunto de entradas de valor contínuo em uma de duas classes. O Perceptron calcula uma soma ponderada das entradas, subtrai um limite e passa um dos dois valores possíveis como resultado. Infelizmente, o Perceptron é limitado e foi

comprovado como tal durante os “anos desiludidos” por Marvin Minsky e o livro de Seymour Papert de 1969, Perceptrons.

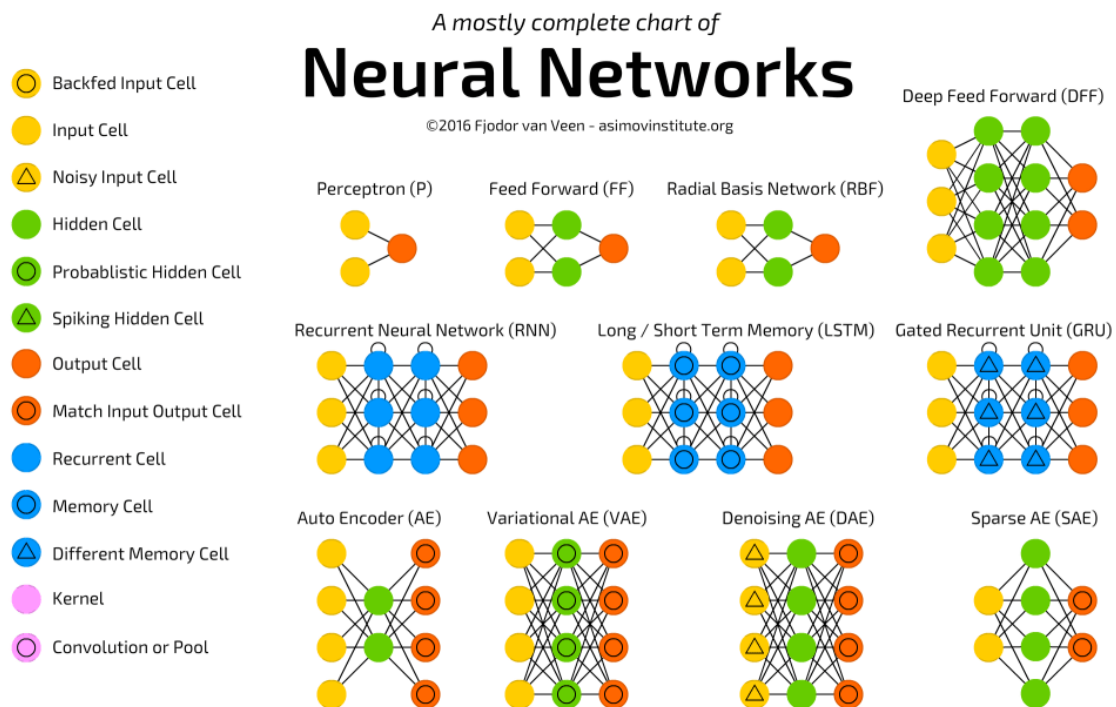


Fig2 – Algumas Arquiteturas de Redes Neurais

Em 1959, Bernard Widrow e Marcian Hoff, de Stanford, desenvolveram modelos denominados “ADALINE” e “MADALINE”. Em uma exibição típica do amor de Stanford por siglas, os nomes provêm do uso de múltiplos elementos ADAPtive LINEar. ADALINE foi desenvolvido para reconhecer padrões binários de modo que, se ele estivesse lendo bits de transmissão de uma linha telefônica, poderia prever o próximo bit. MADALINE foi a primeira rede neural aplicada a um problema do mundo real, usando um filtro adaptativo que elimina ecos nas linhas telefônicas. Embora o sistema seja tão antigo como os sistemas de controle de tráfego aéreo, ele ainda está em uso comercial.

Infelizmente, esses sucessos anteriores levaram as pessoas a exagerar o potencial das redes neurais, particularmente à luz da limitação na eletrônica, então disponível na época. Este exagero excessivo, que decorreu do mundo acadêmico e técnico, infectou a literatura geral da época. Muitas promessas foram feitas, mas o resultado foi o desapontamento. Além disso, muitos escritores começaram a refletir sobre o efeito que teria “máquinas pensantes” no homem. A série de Asimov em robôs revelou os efeitos sobre

a moral e os valores do homem quando máquinas fossem capazes de fazer todo o trabalho da humanidade. Outros escritores criaram computadores mais sinistros, como HAL do filme 2001.

Toda essa discussão sobre o efeito da Inteligência Artificial sobre a vida humana, aliada aos poucos progressos, fizeram vozes respeitadas criticar a pesquisa em redes neurais. O resultado foi a redução drástica de grande parte do financiamento em pesquisas. Esse período de crescimento atrofiado durou até 1981, sendo conhecido como o Inverno da IA (AI Winter).

Em 1982, vários eventos provocaram um renovado interesse. John Hopfield da Caltech apresentou um documento à Academia Nacional de Ciências. A abordagem de Hopfield não era simplesmente modelar cérebros, mas criar dispositivos úteis. Com clareza e análise matemática, ele mostrou como essas redes poderiam funcionar e o que poderiam fazer. No entanto, o maior recurso de Hopfield foi seu carisma. Ele era articulado e simpático e isso colaborou bastante para que ele fosse ouvido.

Em 1985, o Instituto Americano de Física começou o que se tornou uma reunião anual – Redes Neurais para Computação. Em 1987, a primeira Conferência Internacional sobre Redes Neurais do Institute of Electrical and Electronic Engineer's (IEEE) atraiu mais de 1.800 participantes.

Em 1986, com redes neurais de várias camadas nas notícias, o problema era como estender a regra Widrow-Hoff para várias camadas. Três grupos independentes de pesquisadores, dentre os quais David Rumelhart, ex-membro do departamento de psicologia de Stanford, apresentaram ideias semelhantes que agora são chamadas de redes Backpropagation porque distribuem erros de reconhecimento de padrões em toda a rede. As redes híbridas utilizavam apenas duas camadas, essas redes de Backpropagation utilizam muitas. O resultado é que as redes de Backpropagation “aprendem” de forma mais lenta, pois necessitam, possivelmente, de milhares de iterações para aprender, mas geram um resultado muito preciso.

Agora, as redes neurais são usadas em várias aplicações. A ideia fundamental por trás da natureza das redes neurais é que, se ela funcionar na natureza, deve ser capaz de funcionar em computadores. O futuro das redes neurais, no entanto, reside no desenvolvimento de hardware. As redes neurais rápidas e eficientes dependem do hardware especificado para seu eventual uso.

O diagrama abaixo mostra alguns marcos importantes na evolução e pesquisa das redes neurais artificiais. O fato, é que ainda estamos escrevendo esta história e muita evolução está ocorrendo neste momento,

através do trabalho de milhares de pesquisadores e profissionais de Inteligência Artificial em todo mundo. E você, não quer ajudar a escrever esta história?

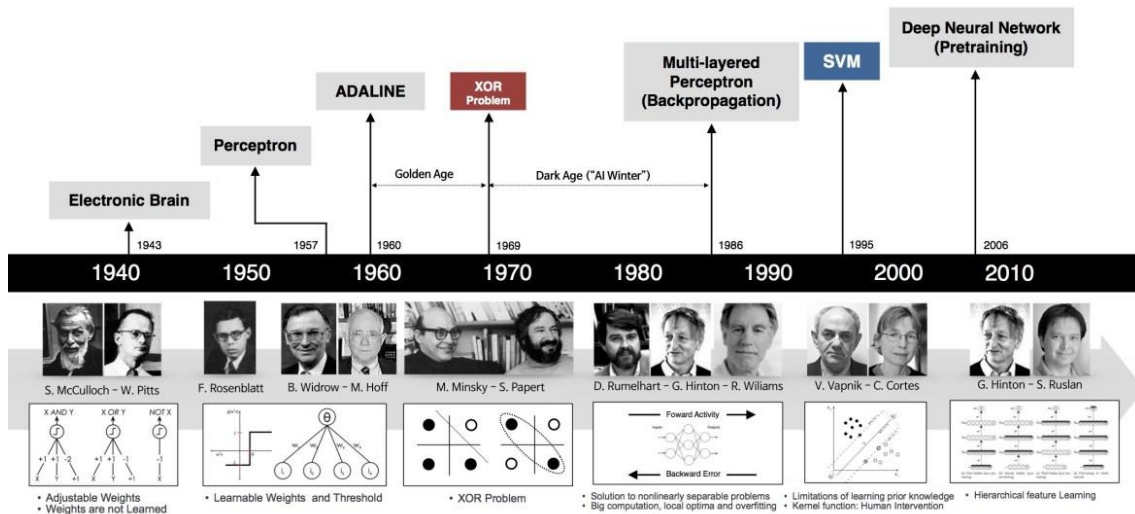


Fig3 – Marcos no desenvolvimento das redes neurais.

Podemos resumir assim os principais marcos na pesquisa e evolução das redes neurais artificiais até chegarmos ao Deep Learning:

1943: Warren McCulloch e Walter Pitts criam um modelo computacional para redes neurais baseadas em matemática e algoritmos denominados lógica de limiar.

1958: Frank Rosenblatt cria o Perceptron, um algoritmo para o reconhecimento de padrões baseado em uma rede neural computacional de duas camadas usando simples adição e subtração. Ele também propôs camadas adicionais com notações matemáticas, mas isso não seria realizado até 1975.

1980: Kunihiro Fukushima propõe a Neoconitron, uma rede neural de hierarquia, multicamada, que foi utilizada para o reconhecimento de caligrafia e outros problemas de reconhecimento de padrões.

1989: os cientistas conseguiram criar algoritmos que usavam redes neurais profundas, mas os tempos de treinamento para os sistemas foram medidos em dias, tornando-os impraticáveis para o uso no mundo real.

1992: Juyang Weng publica o Cresceptron, um método para realizar o reconhecimento de objetos 3-D automaticamente a partir de cenas desordenadas.

Meados dos anos 2000: o termo “aprendizagem profunda” começa a ganhar popularidade após um artigo de Geoffrey Hinton e Ruslan Salakhutdinov mostrar como uma rede neural de várias camadas poderia ser pré-treinada uma camada por vez.

2009: acontece o NIPS Workshop sobre Aprendizagem Profunda para Reconhecimento de Voz e descobre-se que com um conjunto de dados suficientemente grande, as redes neurais não precisam de pré-treinamento e as taxas de erro caem significativamente.

2012: algoritmos de reconhecimento de padrões artificiais alcançam desempenho em nível humano em determinadas tarefas. E o algoritmo de aprendizagem profunda do Google é capaz de identificar gatos.

2014: o Google compra a Startup de Inteligência Artificial chamada DeepMind, do Reino Unido, por £ 400m

2015: Facebook coloca a tecnologia de aprendizado profundo – chamada DeepFace – em operação para marcar e identificar automaticamente usuários do Facebook em fotografias. Algoritmos executam tarefas superiores de reconhecimento facial usando redes profundas que levam em conta 120 milhões de parâmetros.

2016: o algoritmo do Google DeepMind, AlphaGo, mapeia a arte do complexo jogo de tabuleiro Go e vence o campeão mundial de Go, Lee Sedol, em um torneio altamente divulgado em Seul.

2017: adoção em massa do Deep Learning em diversas aplicações corporativas e mobile, além do avanço em pesquisas. Todos os eventos de tecnologia ligados a Data Science, IA e Big Data, apontam Deep Learning como a principal tecnologia para criação de sistemas inteligentes.

A promessa do aprendizado profundo não é que os computadores comecem a pensar como seres humanos. Isso é como pedir uma maçã para se tornar uma laranja. Em vez disso, demonstra que, dado um conjunto de dados suficientemente grande, processadores rápidos e um algoritmo suficientemente sofisticado, os computadores podem começar a realizar tarefas que até então só podiam ser realizadas apenas por seres humanos, como reconhecer imagens e voz, criar obras de arte ou tomar decisões por si mesmo.

Os estudos sobre as redes neurais sofreram uma grande revolução a partir dos anos 80 e esta área de estudos tem se destacado, seja pelas promissoras

características apresentadas pelos modelos de redes neurais propostos, seja pelas condições tecnológicas atuais de implementação que permitem desenvolver arrojadas implementações de arquiteturas neurais paralelas em hardwares dedicado, obtendo assim ótimas performances destes sistemas (bastante superiores aos sistemas convencionais). A evolução natural das redes neurais, são as redes neurais profundas (ou Deep Learning). Mas isso é o que vamos discutir no próximo capítulo! Até lá.

Referências:

Christopher D. Manning. (2015). Computational Linguistics and Deep Learning Computational Linguistics, 41(4), 701–707.

F. Rosenblatt. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133, 1943.

The organization of behavior: A neuropsychological theory. D. O. Hebb. John Wiley And Sons, Inc., New York, 1949

B. Widrow et al. Adaptive "Adaline" neuron using chemical "memistors". Number Technical Report 1553-2. Stanford Electron. Labs., Stanford, CA, October 1960.

"New Navy Device Learns By Doing", New York Times, July 8, 1958.

Perceptrons. An Introduction to Computational Geometry. MARVIN MINSKY and SEYMOUR PAPERT. M.I.T. Press, Cambridge, Mass., 1969.

Minsky, M. (1952). A neural-analogue calculator based upon a probability model of reinforcement. Harvard University Psychological Laboratories internal report.

Capítulo 3 – O Que São Redes Neurais Artificiais Profundas ou Deep Learning?

veremos neste capítulo. Não se preocupe se alguns termos mais técnicos não fizerem sentido agora. Todos eles serão estudados ao longo deste livro online.

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.

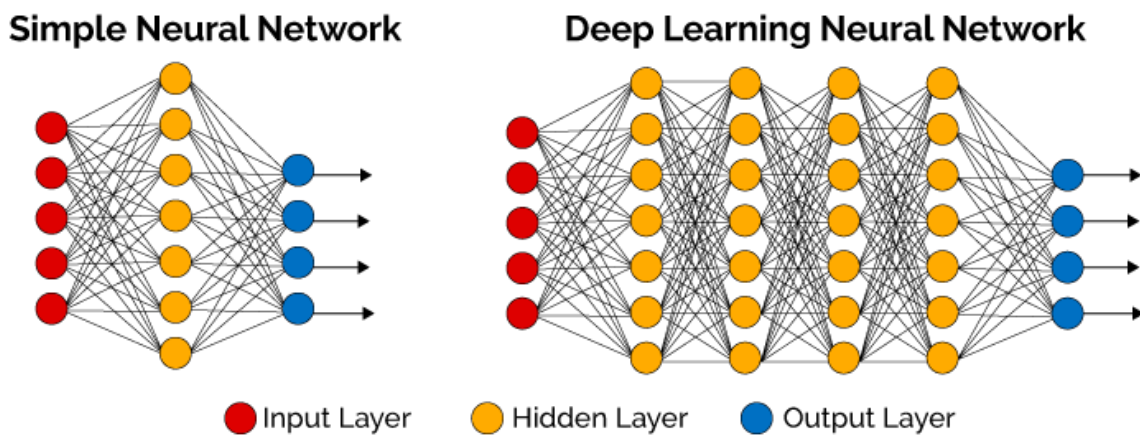


Fig4 – Rede Neural Simples e Rede Neural Profunda (Deep Learning)

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 surpreendente 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 (Figura 5), 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).

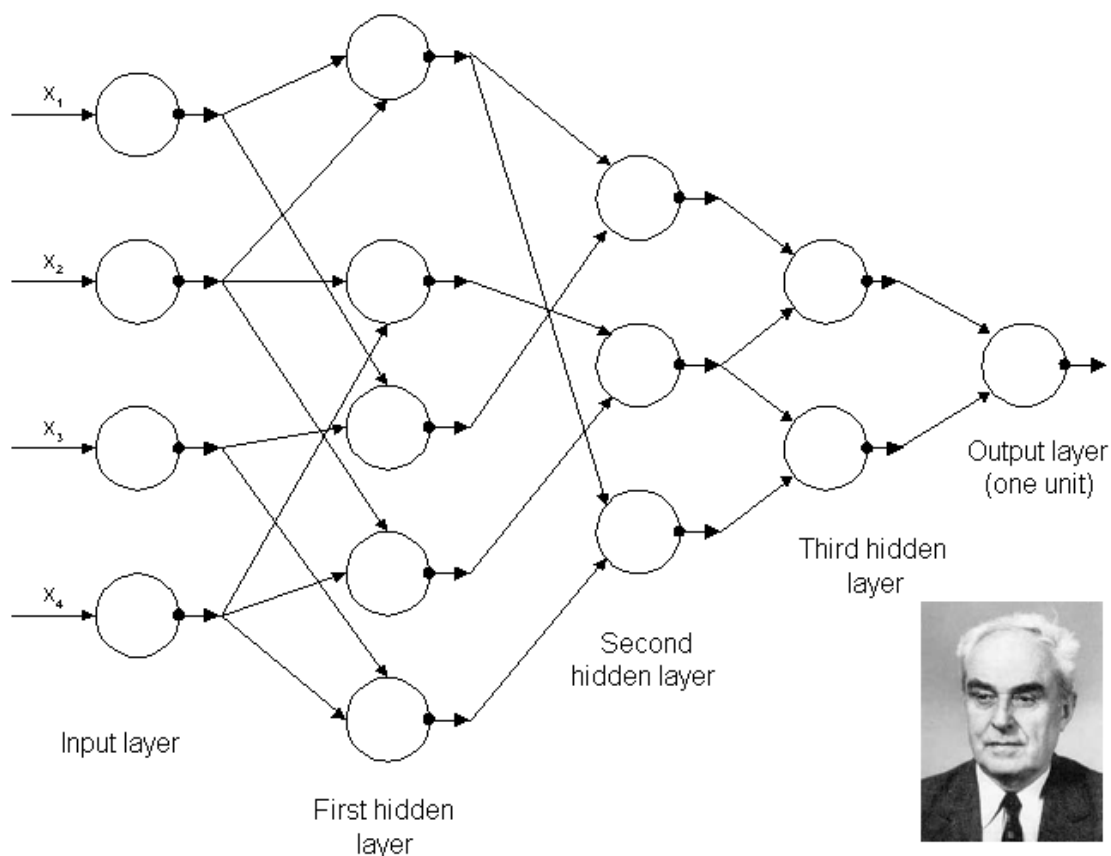


Fig5 – Arquitetura da primeira rede profunda conhecida treinada por Alexey Grigorevich Ivakhnenko em 1965.

No final da década de 1970, o primeiro inverno de AI 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 Kuniyuki 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.



Fig6 – Os pioneiros 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.

No próximo capítulo você vai começar a compreender tecnicamente como funciona a Aprendizagem Profunda. Até o capítulo 4.

Referências:

Deep Learning in a Nutshell: History and Training from NVIDIA

Linnainmaa, S. (1970). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's thesis, Univ. Helsinki.

P. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, Cambridge, MA, 1974.

Werbos, P.J. (2006). Backwards differentiation in AD and neural nets: Past links and new opportunities. In Automatic Differentiation: Applications, Theory, and Implementations, pages 15-34. Springer.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.

Widrow, B., & Lehr, M. (1990). 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 78(9), 1415-1442.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. 1986. Learning internal representations by error propagation. In *Parallel distributed processing: explorations in the microstructure of cognition*, vol. 1, David E. Rumelhart, James L. McClelland, and CORPORATE PDP Research Group (Eds.). MIT Press, Cambridge, MA, USA 318-362