

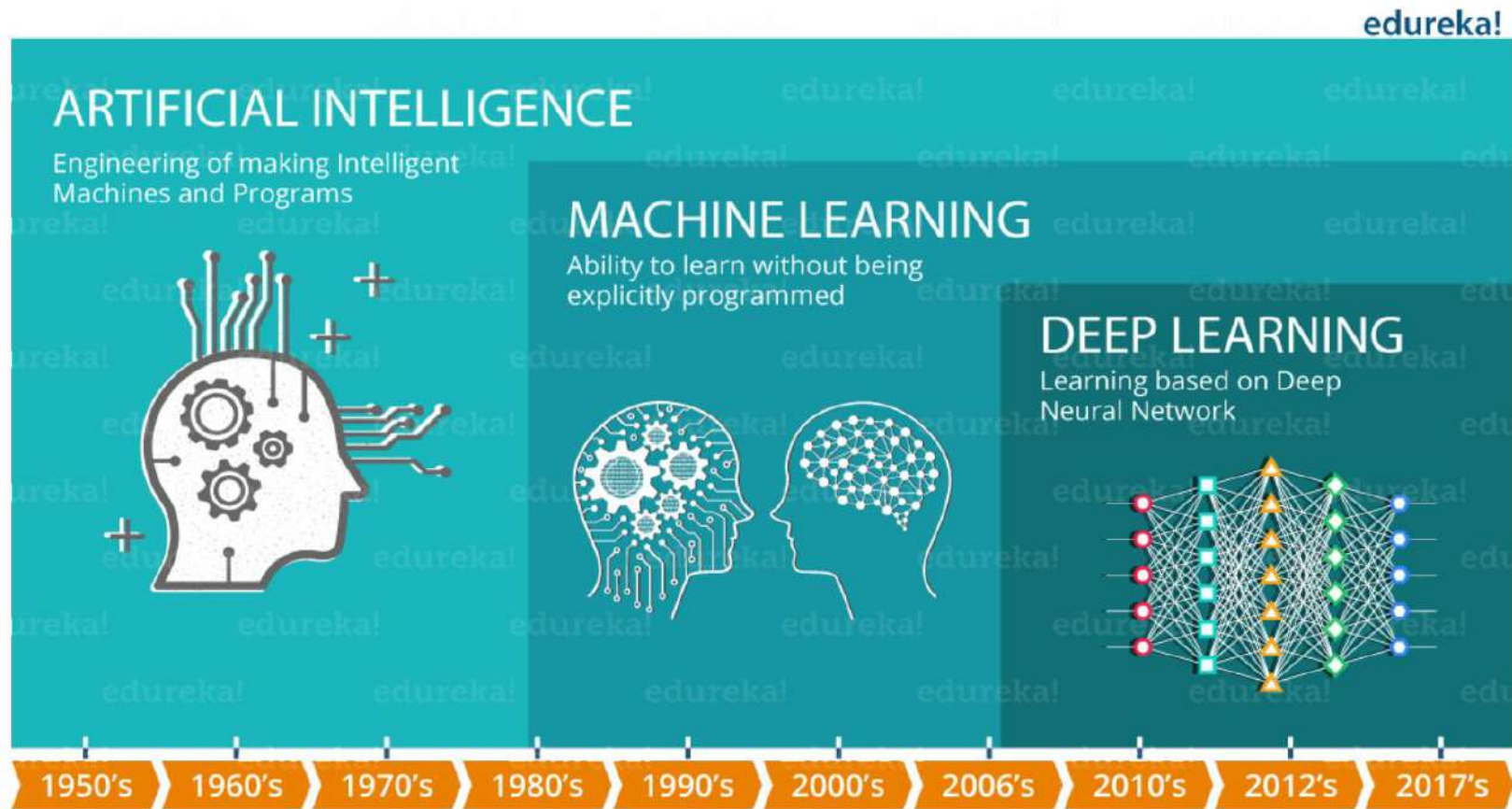
# IA353 – Redes Neurais

## Tópico 1 – Introdução e Motivação

### Índice Geral

1	Visão geral de aprendizado de máquina.....	2
2	Principais módulos do aprendizado de máquina .....	7
3	Principais vertentes do aprendizado de máquina .....	8
4	Regularização e capacidade de generalização.....	12
5	Um exemplo didático em treinamento supervisionado .....	19
6	A inteligência artificial já é uma realidade.....	48
7	Computação Digital e Neurocomputação .....	52
8	Lei de Moore e a expansão tecnológica.....	56
9	O poder da simulação computacional .....	69
10	Ser igual × Ser não-distinguível .....	72
	10.1 Etapas do processo de aprendizado.....	78
	10.2 A sala chinesa de Searle .....	81
	10.3 O teste de Turing .....	83
11	Uma excursão por <i>deep learning</i> .....	87
12	Referências bibliográficas .....	101

# 1 Visão geral de aprendizado de máquina



Fonte: <https://www.edureka.co/blog/what-is-deep-learning>



Figura 1 – Perspectiva crítica para o aprendizado de máquina

Fonte: <https://xkcd.com/1838/>

- O aprendizado de máquina envolve a extração de conhecimento diretamente de dados disponíveis, visando “modelar a porção do mundo” de interesse, realizar previsões e dar suporte à tomada de decisão, dentre outras aplicações.
- Ele está fundamentado em teoria de probabilidade, estatística e otimização, sustenta técnicas indispensáveis em soluções para *big data*, ciência de dados, reconhecimento de padrões e recuperação de informação, além de fomentar os principais progressos recentes em inteligência artificial.
- Em termos de aplicações práticas, técnicas de aprendizado de máquina se tornaram ingredientes críticos em reconhecimento de padrões, visão computacional, processamento de linguagem natural e robótica.

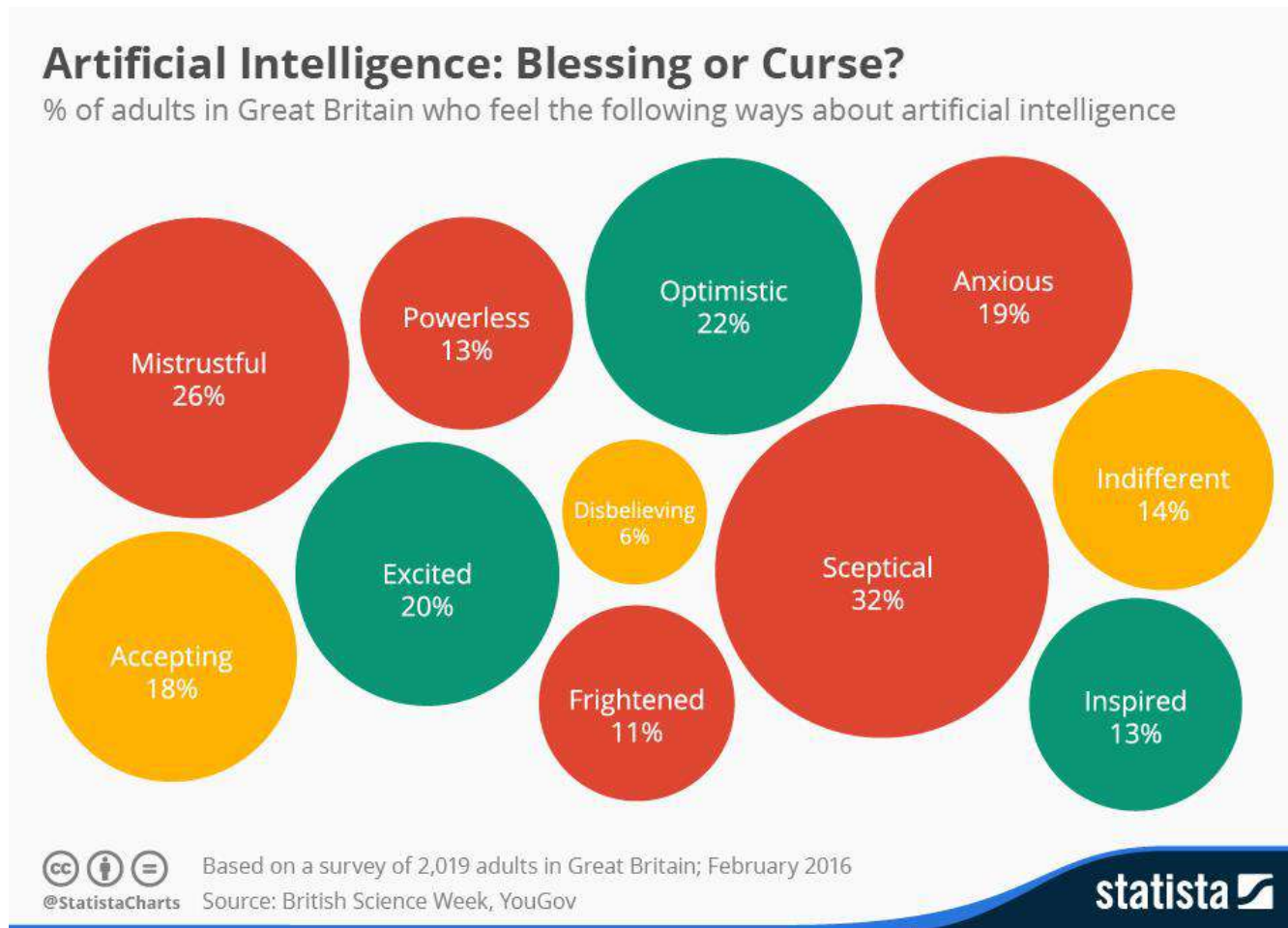


Figura 2 – Sensações despertadas nas pessoas por parte de soluções de IA

Fonte (Março, 2016): <https://www.statista.com/chart/4503/views-on-artificial-intelligence/>

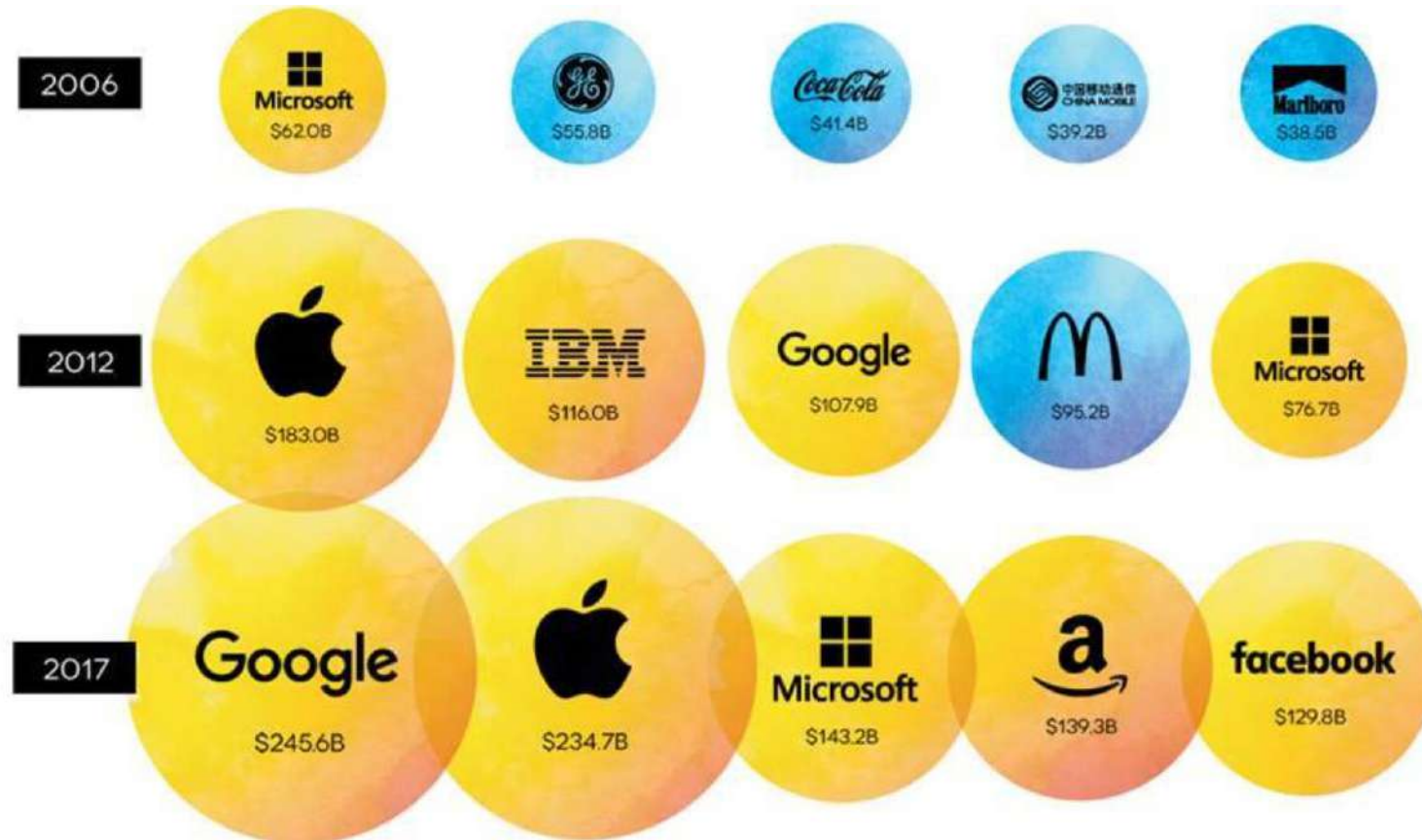


Figura 3 – Evolução da lista das 5 empresas com maior valor de marca perante os consumidores

Fonte (August, 2018): <https://venturebeat.com/2018/08/19/how-to-create-a-virtuous-cycle-of-data-with-your-customers/>

## 2 Principais módulos do aprendizado de máquina

- Em termos formais, técnicas de aprendizado de máquina são compostas por:
  - Um **conjunto de dados**: pode ser particionado nos conjuntos de treinamento, validação e teste, que serão melhor descritos mais adiante.
  - Um **modelo matemático** (modelo de aprendizado) para a síntese de um mapeamento multidimensional de entrada-saída, que possui algum grau de plasticidade ou flexibilidade a partir da presença de parâmetros livres.
  - Uma **função de perda ou de custo** que mede o desempenho do modelo de aprendizado: por exemplo, pode ser o erro quadrático médio em problemas de regressão e a taxa de erro de classificação em problemas de classificação.
  - Um **procedimento de otimização** de parâmetros livres do modelo visando minimizar as perdas ou o custo ao se caminhar pela superfície de erro, que pode ou não ser convexa.

### 3 Principais vertentes do aprendizado de máquina

- Da seção anterior, fica evidente que uma máquina aprende quando realiza o ajuste de parâmetros livres do modelo de aprendizado. No caso de redes neurais artificiais como modelos de aprendizado, os parâmetros livres são os pesos das conexões sinápticas.
- Usualmente, a área de aprendizado de máquina é dividida em: aprendizado supervisionado, aprendizado não-supervisionado e aprendizado por reforço.
- Em **aprendizado supervisionado**, há uma quantidade finita e possivelmente ruidosa de dados rotulados, indicando (com limitações) o comportamento desejado para a máquina de aprendizado. Quando as saídas desejadas são discretas ou categóricas, geralmente se fala de **problemas de classificação** supervisionada. Já para o caso de saídas contínuas ou numéricas, tem-se **problemas de regressão**. De qualquer modo, ambos envolvem a aproximação de mapeamentos



multidimensionais contínuos de entrada-saída, com múltiplos estímulos de entrada e múltiplas saídas desejadas (MIMO: *Multiple Input, Multiple Output*).

- Técnicas de **aprendizado não-supervisionado** buscam extrair conhecimento a partir de dados não-rotulados, por exemplo, realizando clusterização e estimação de densidade (lei de distribuição que responde pela geração dos dados). A clusterização visa agrupar, ou seja, classificar numa mesma classe, objetos que apresentem atributos similares (requer o emprego de métricas de distância), mantendo em grupos distintos objetos que apresentem atributos dissimilares. A estimação de densidade procura aproximar a função densidade de probabilidade que supostamente gerou os dados, maximizando uma função de verossimilhança que torna os dados mais prováveis sob este modelo estatístico.
- Cabe interromper momentaneamente as definições das vertentes do aprendizado de máquina para introduzir o conceito de **aprendizado da representação**, tema muito relevante para o estudo e a síntese de modelos de *deep learning*. É

predominantemente associado a técnicas de aprendizado não-supervisionado. O objetivo é encontrar uma ou mais representações (em cascata e geralmente não-lineares) que preservem o máximo possível de informação relevante dos dados originais, ao mesmo tempo em que conduzem a uma configuração mais simples e acessível dos dados, quando comparada com a configuração original. Configurações mais simples e acessíveis são geralmente associadas a representações em espaços de menor dimensão (embora haja opções interessantes que expandem a dimensão do espaço), representações que promovem esparsidade e representações que produzem maior “ortogonalidade” entre os atributos. Aprender a representação é equivalente a sintetizar um elenco consistente de **filtros extratores de atributos relevantes da entrada**.

- Em **aprendizado por reforço**, não há um claro indicativo do que é certo e do que é errado o tempo todo, como ocorre em aprendizado supervisionado, mas esporadicamente é recebido um grau de avaliação de desempenho (recompensa ou

punição) de uma tarefa caracterizada por uma sequência de decisões. O objetivo então é maximizar a recompensa acumulada a longo prazo, ou seja, encontrar a melhor estratégia de tomada de decisões sequenciais, também chamada de política ótima. Trata-se de uma técnica que geralmente recorre a ambientes simulados em computador. É impressionante o sucesso do aprendizado por reforço, por exemplo, em jogos de tabuleiro, particularmente pelo fato da máquina poder jogar contra si própria enquanto aprende. Lembre-se que o tempo de uma simulação computacional em um ambiente virtual pode ser muito acelerado em relação ao tempo físico de ocorrência do experimento no mundo real.

- Redes neurais com arquiteturas profundas, que requerem *deep learning*, estão entre as mais bem-sucedidas vertentes de aprendizado de máquina e se inserem como módulos decisivos em sistemas caracterizados por aprendizado supervisionado, não-supervisionado e por reforço.

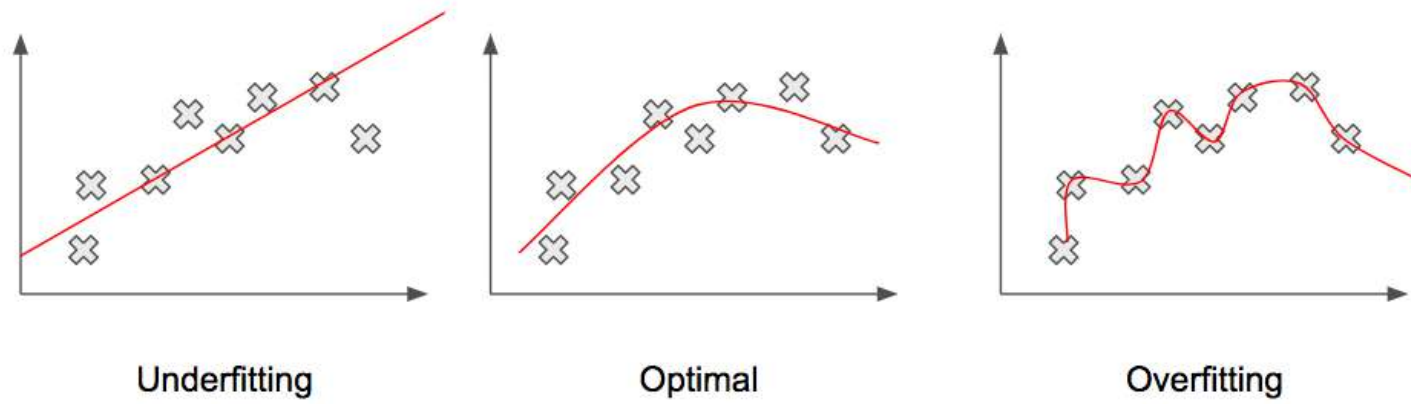
## 4 Regularização e capacidade de generalização

- Embora não esteja restrito a este caso, esta seção vai se concentrar em técnicas de aprendizado supervisionado.
- Os modelos de aprendizado geralmente são tomados como sendo mais flexíveis do que a aplicação vai demandar. De fato, boa parte dos modelos de aprendizado mais populares possuem a **capacidade de aproximação universal**, propriedade fundamental quando não se tem ideia da natureza do mapeamento demandado.
- Logo, são dois os desafios durante o treinamento: (1) obter o melhor proveito do modelo de aprendizado, definindo adequadamente a sua **flexibilidade** (processo também chamado de **regularização do modelo de aprendizado**), geralmente a partir de **hiperparâmetros** associados à formulação matemática do modelo; (2) obter o melhor proveito dos dados disponíveis, treinando na medida certa, o que também vai requer outros **hiperparâmetros** adequadamente definidos, particularmente junto ao algoritmo de otimização.

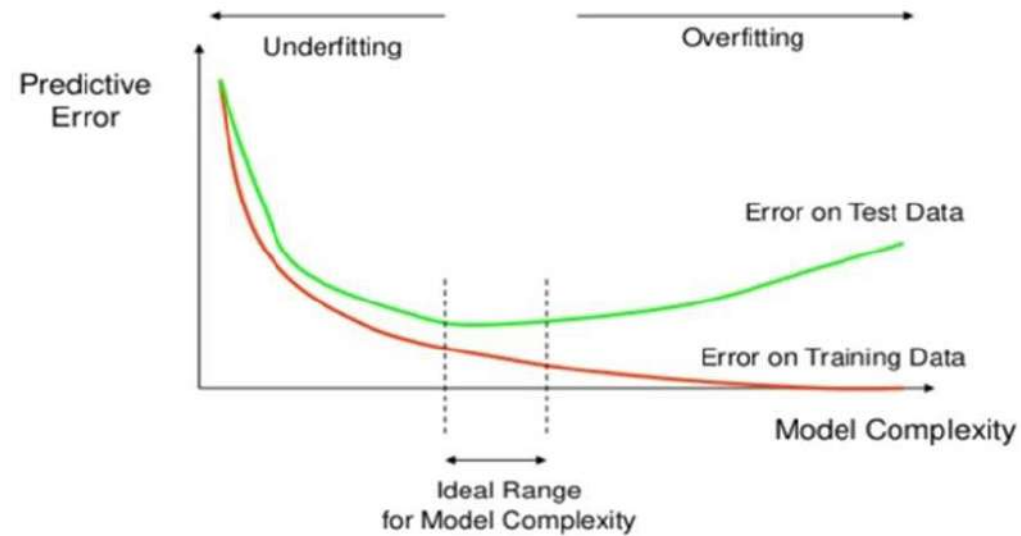
- Se esses dois desafios forem bem trabalhados, diz-se que a máquina, ao final do processo de aprendizado, possui **máxima capacidade de generalização**, ou seja, exibe o melhor desempenho para dados não vistos durante o treinamento.
- Esses dados não vistos durante o treinamento são chamados de dados de teste.
- Dessa forma, procedimentos e métodos gerais em aprendizado de máquina, particularmente no contexto de aprendizado supervisionado, incluem as etapas de **pré-processamento e partição dos dados disponíveis** e de **sintonia de hiperparâmetros** dos modelos de aprendizado e das técnicas de otimização.
- Muitos aspectos envolvidos no treinamento de redes neurais artificiais com camadas profundas (*deep learning*) não contam com uma fundamentação teórica adequada, embora progressos têm sido verificados nesses últimos anos. Questões que já começam a ser apropriadamente respondidas são:
  1. Em que circunstâncias vai funcionar bem?
  2. Como implementar o auto-ajuste de hiperparâmetros associados ao algoritmo de otimização que implementa o processo de treinamento?

3. Como definir otimamente hiperparâmetros como número de camadas, tipos de camadas, número de neurônios por camada e tipos de neurônios?

- Para essas questões, um procedimento válido ainda é tentativa-e-erro, sendo que casos de sucesso na literatura geralmente balizam novas aplicações.
- Iremos abordar neste curso técnicas de busca automática que respondem de forma mais eficaz aos itens da questão (3).
- A **capacidade de generalização** de uma máquina de aprendizado pode ser adequadamente medida pelo erro junto aos dados de validação, durante o treinamento, e junto aos dados de teste, após o término do treinamento. Essencialmente, os parâmetros livres são ajustados visando reduzir o erro junto aos dados de treinamento, de forma iterativa, sendo que o treinamento é interrompido quando se minimiza o erro junto aos dados de validação (não usados no ajuste dos parâmetros livres). Parar o treinamento antes do mínimo erro de validação é considerado sub-ajuste (do inglês *underfitting*), enquanto prosseguir no treinamento após este mínimo é considerado sobreajuste (do inglês *overfitting*).

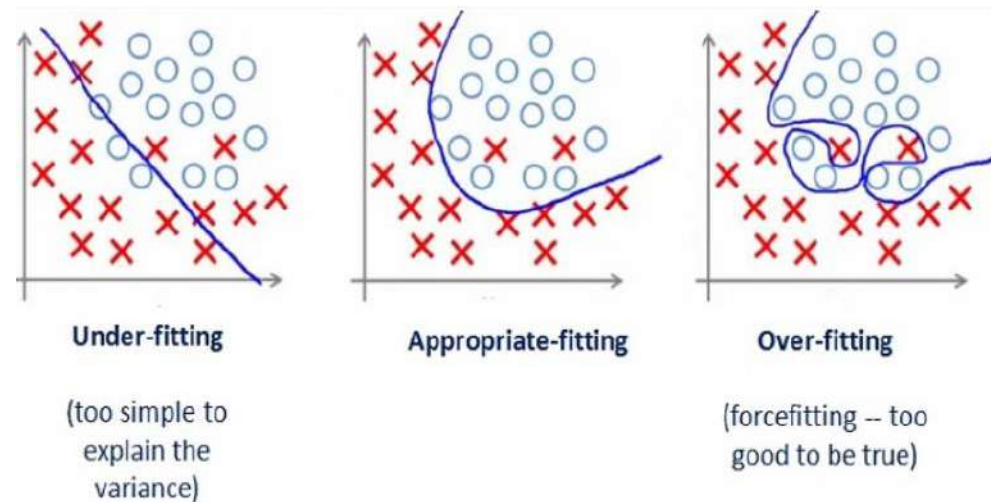


Fonte: <https://pythonmachinelearning.pro/a-guide-to-improving-deep-learnings-performance/>

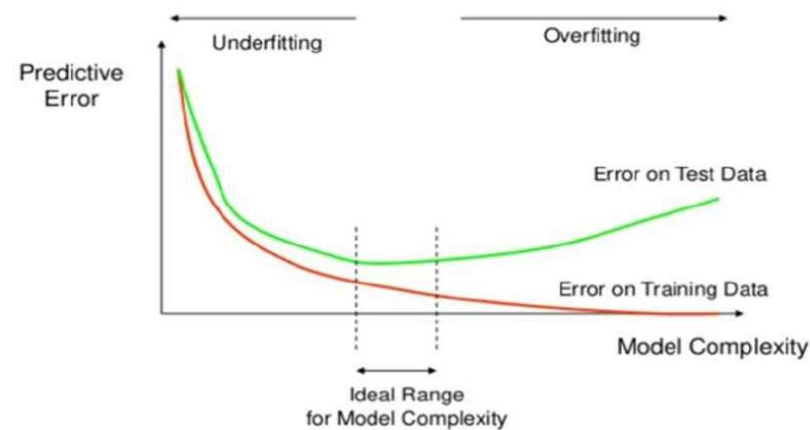


Fonte: <https://stats.stackexchange.com/questions/292283/general-question-regarding-over-fitting-vs-complexity-of-models>

- Supõe-se aqui que, quanto mais se treina um modelo de aprendizado, mais se explora sua flexibilidade, resultando num mapeamento mais complexo.



Fonte: <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>





- Este procedimento é eficaz quando os conjuntos de treinamento, validação e teste são **independentes e suficientemente representativos do processo que se quer aprender**.
- Serão vistas outras formas de se buscar maximizar a capacidade de generalização, com ênfase na adição de um termo de regularização que penaliza a flexibilidade do modelo de aprendizado.
- Em *deep learning*, maximizar a capacidade de generalização é bastante desafiador e não depende apenas dos parâmetros livres, uma vez proposta a arquitetura de rede neural. Hiperparâmetros associados aos aspectos estruturais da arquitetura e hiperparâmetros associados ao algoritmo de otimização devem ser adequadamente definidos.
- A técnica ilustrada acima para evitar *overfitting* é denominada **holdout**. Uma forma estatisticamente mais robusta é denominada ***k-fold cross-validation***, sendo caracterizada por uma melhor exploração do conjunto de dados.

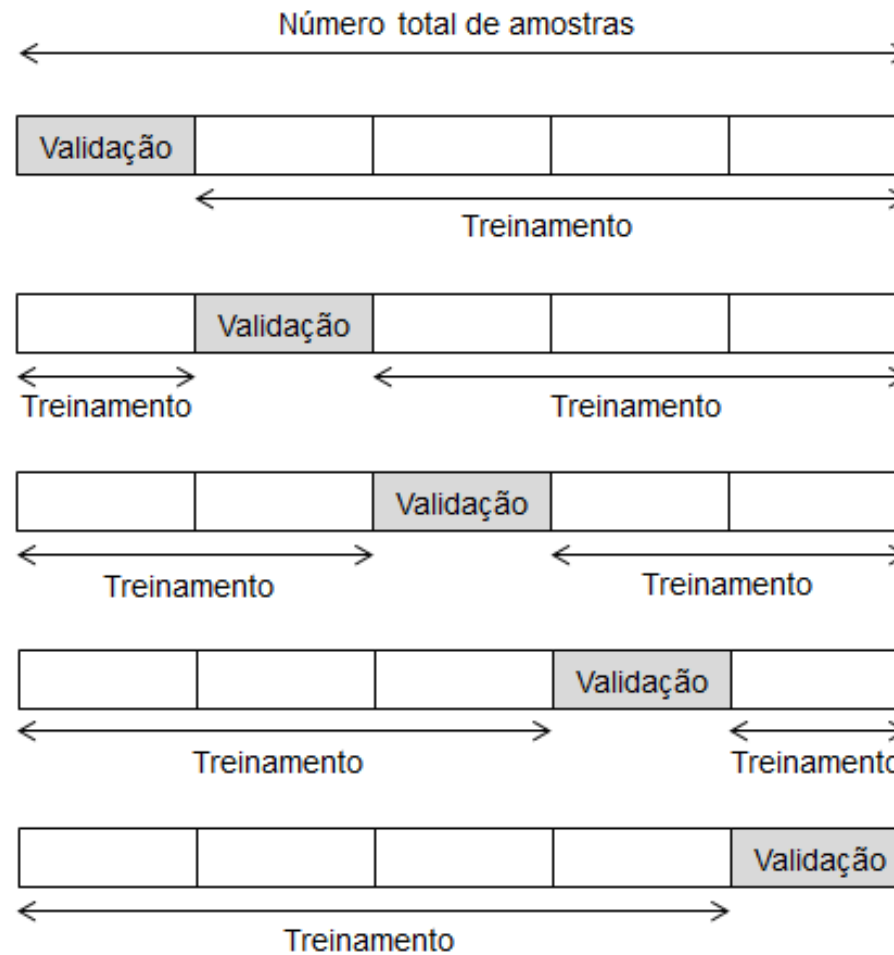


Figura 4 – Esquema de divisão das amostras disponíveis para treinamento na abordagem de validação cruzada com  $k$ -pastas ( $k$ -fold cross-validation), com  $k=5$ .

## 5 Um exemplo didático em treinamento supervisionado

- Parte deste material foi extraído da palestra de James Mickens (Harvard University), num evento de segurança computacional. Fonte: <https://www.youtube.com/watch?v=ajGX7odA87k>

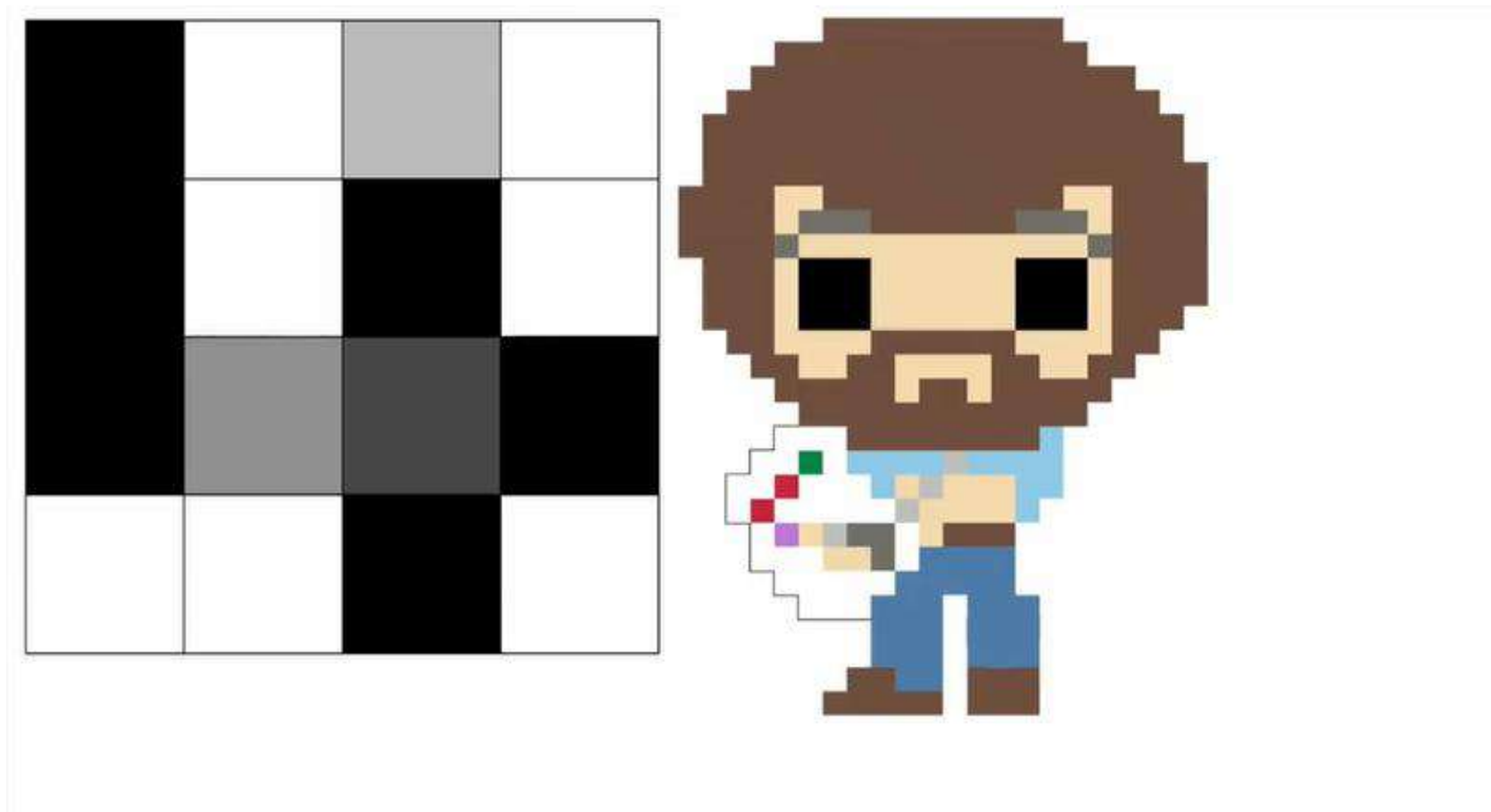


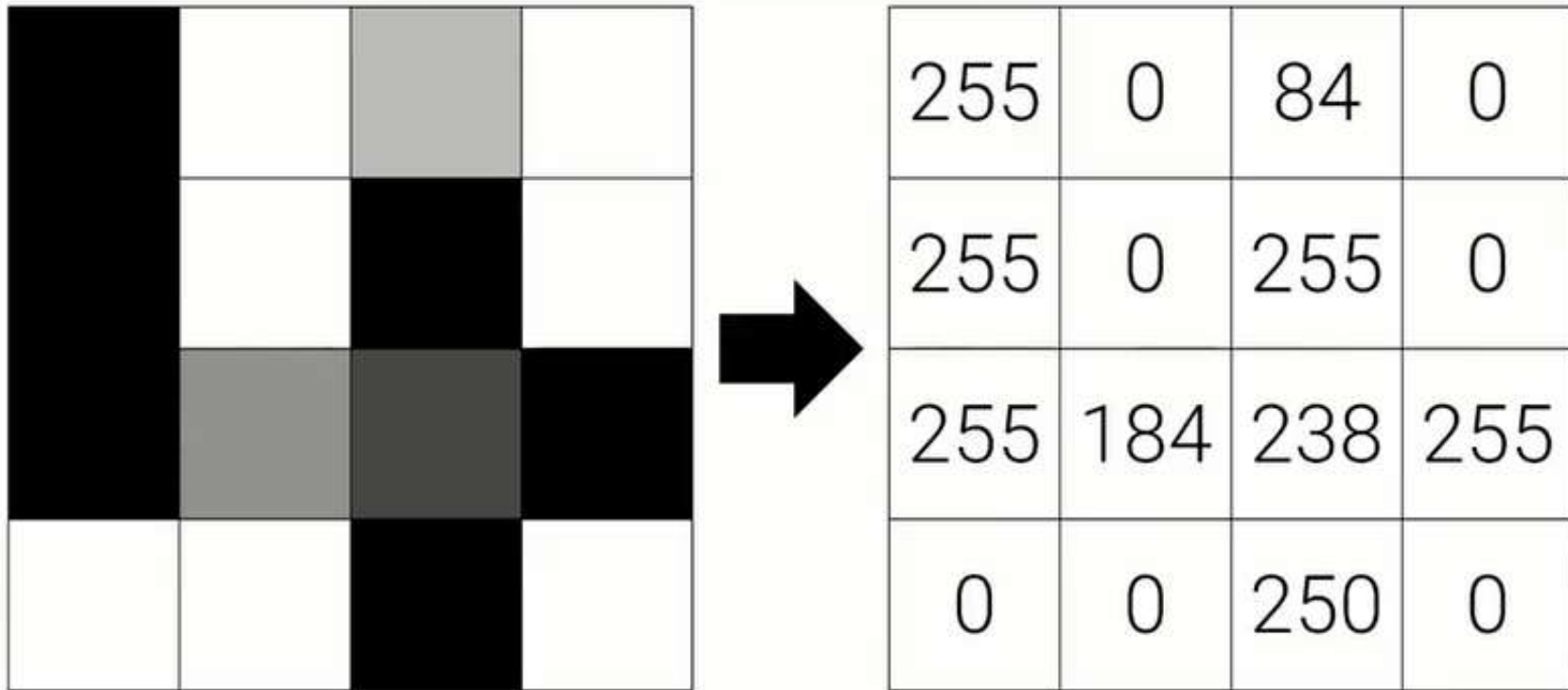
## **Goal of computer security**

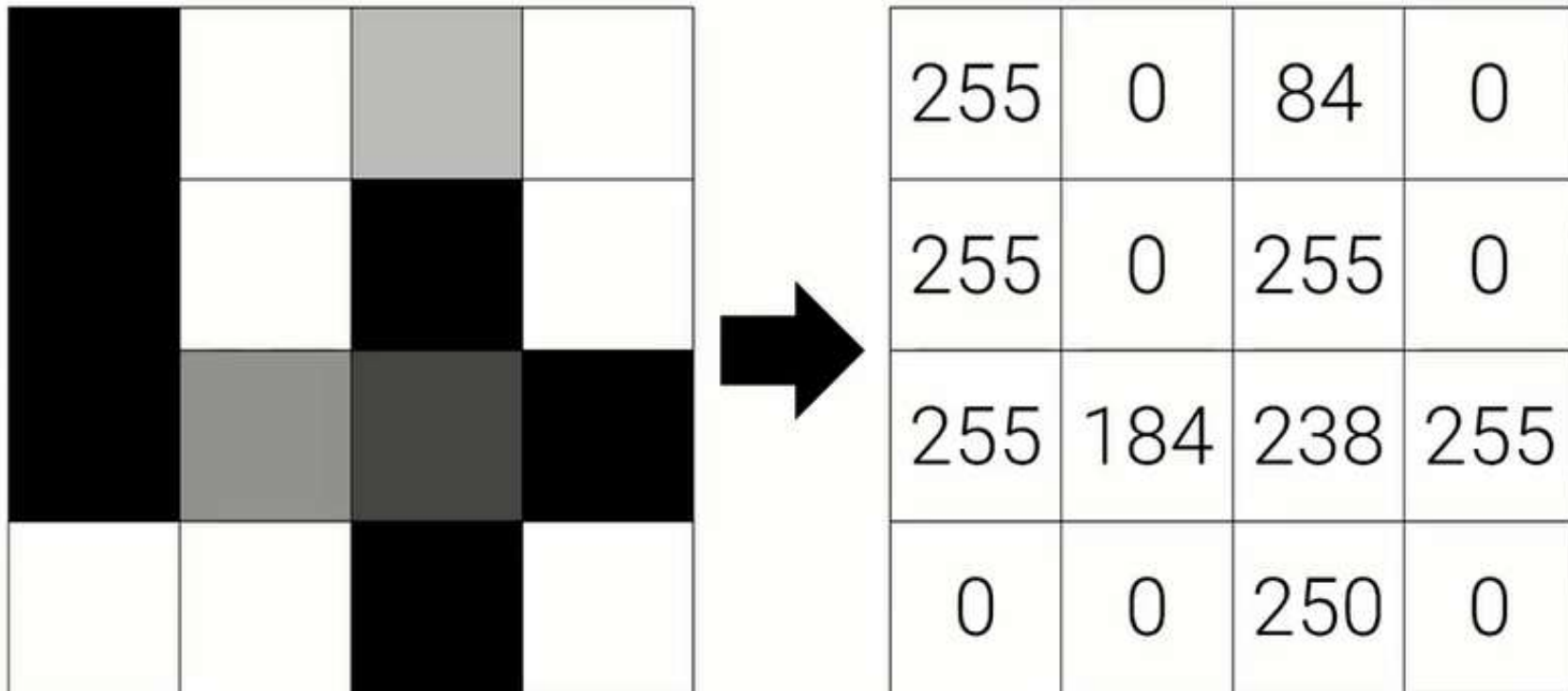
Ensure that systems do the right thing,  
even in the presence of malicious inputs

## **Challenges presented by M.L.**

- (1) Lack of deep understanding about  
how M.L. works
- (2) Many people don't care about (1)



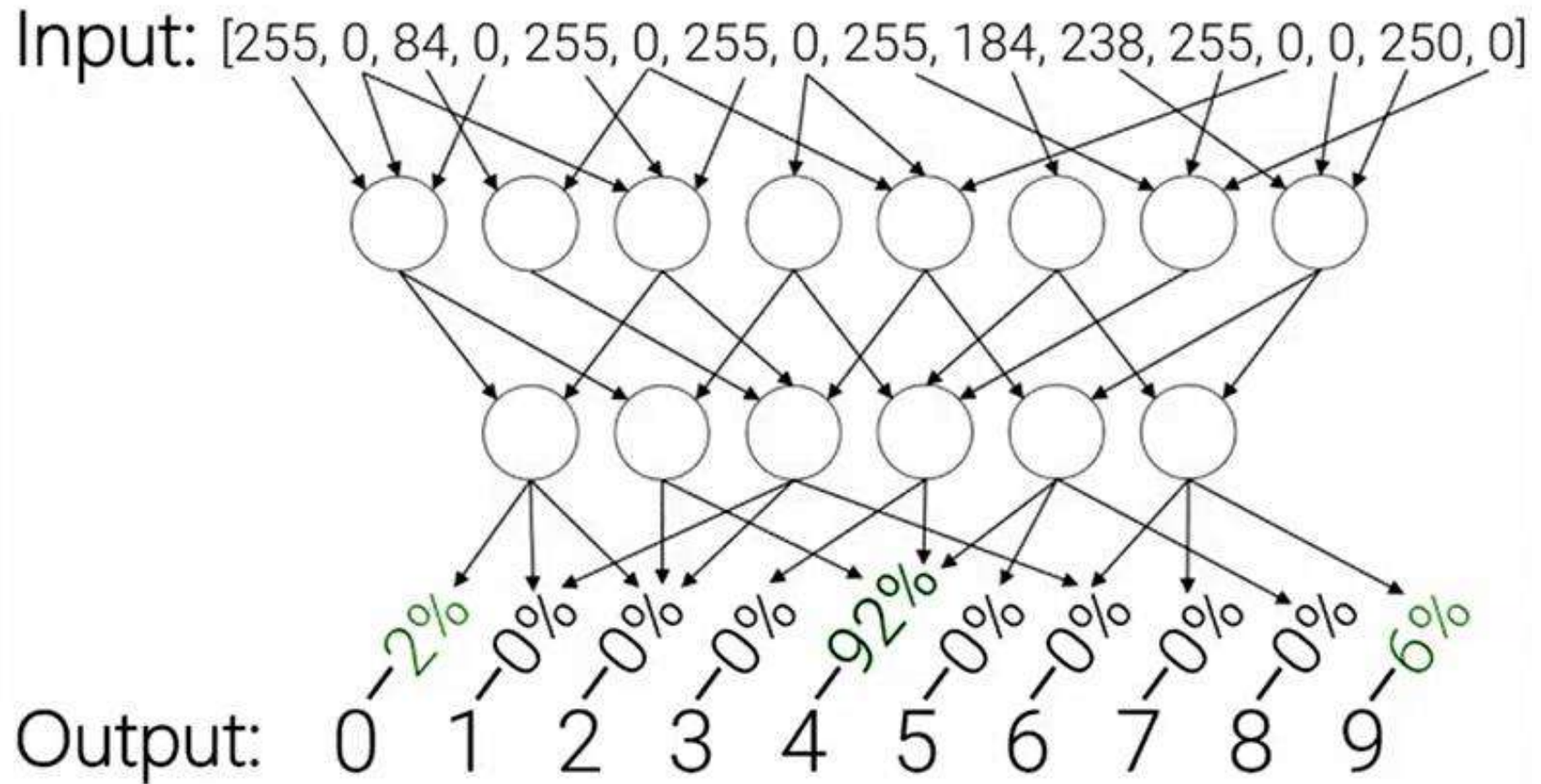


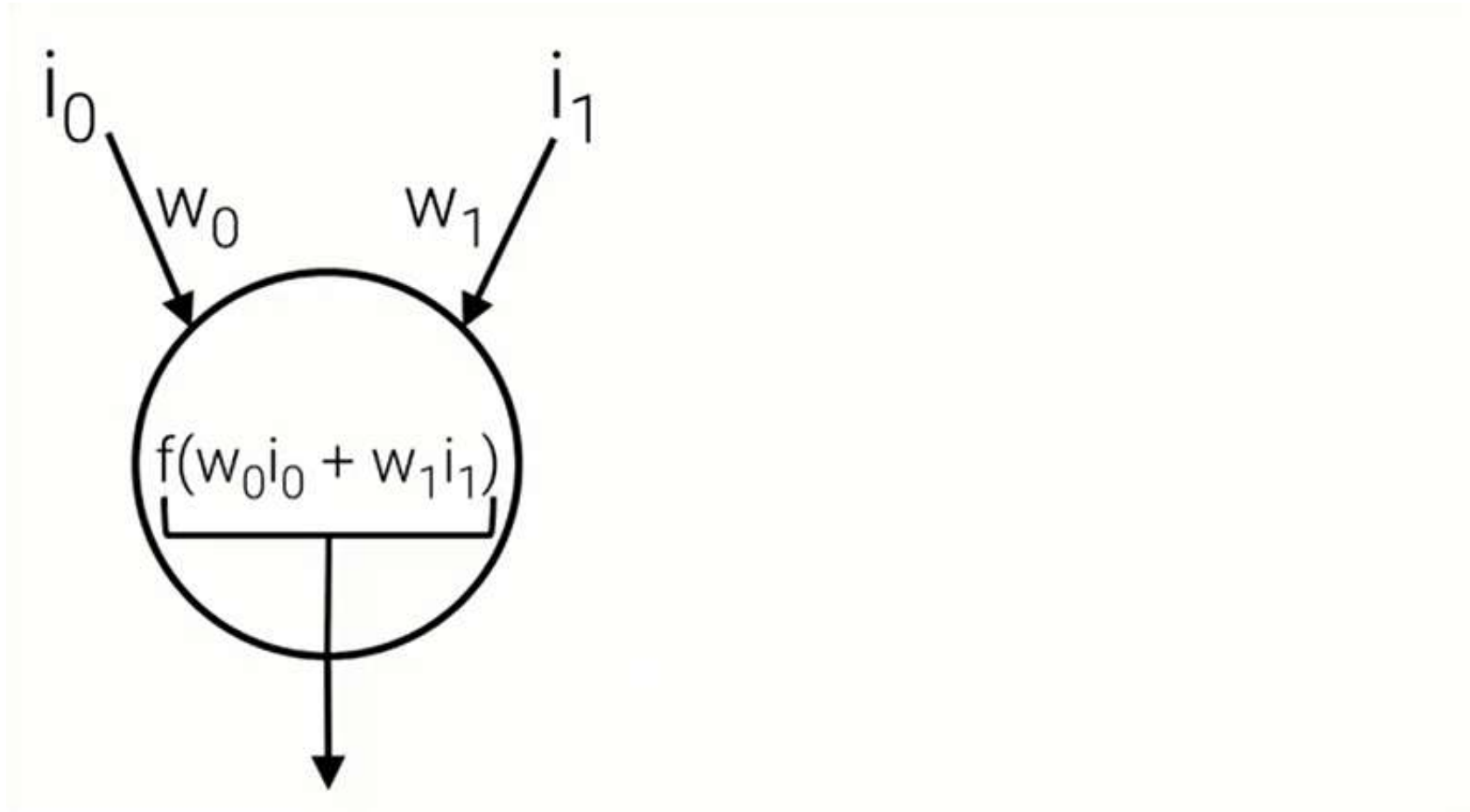
[255, 0, 84, 0, 255, 0, 255, 0, 255, 184, 238, 255, 0, 0, 250, 0]

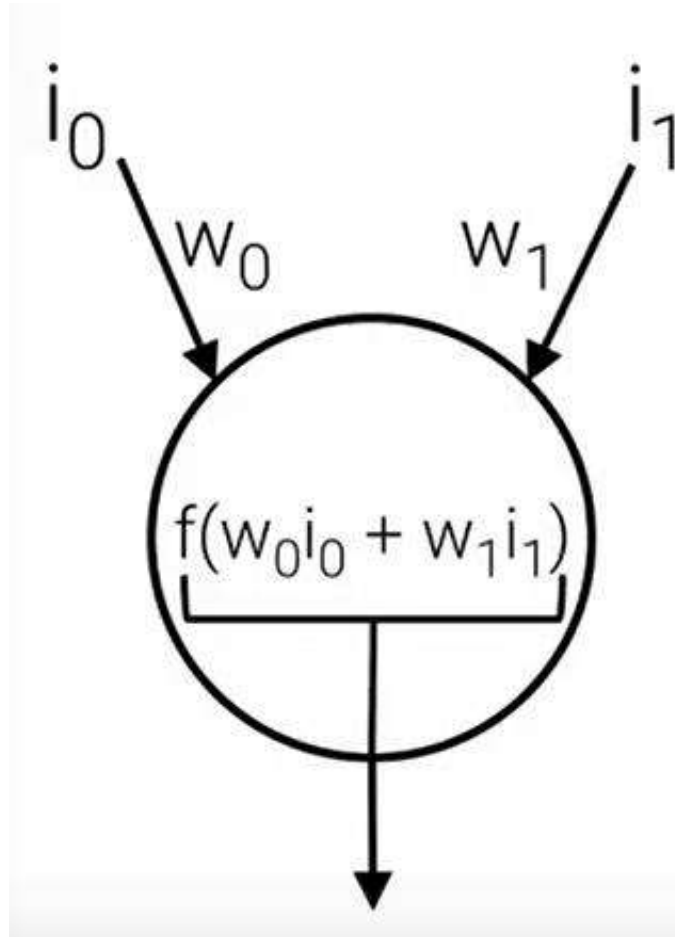


Input: [255, 0, 84, 0, 255, 0, 255, 0, 255, 184, 238, 255, 0, 0, 250, 0]

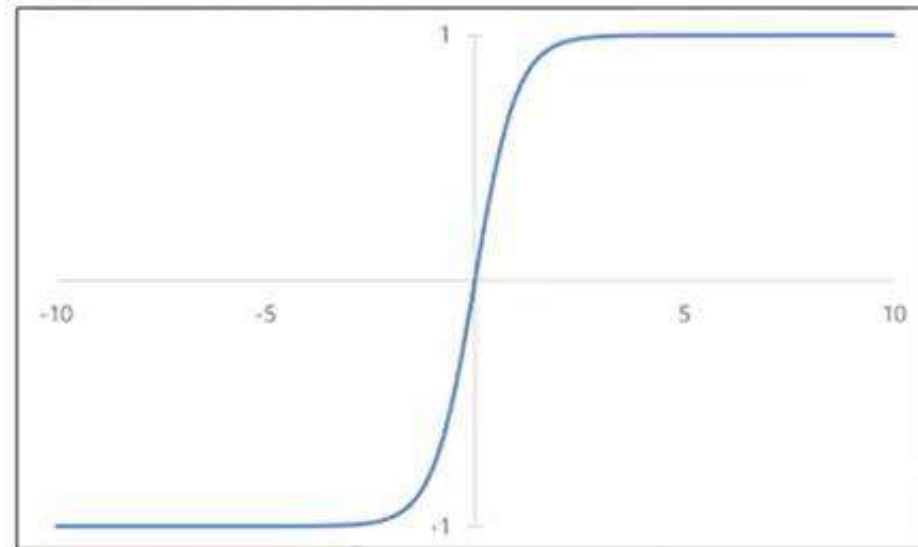
Output: 0 1 2 3 4 5 6 7 8 9  
          2% 0% 0% 0% 92% 0% 0% 0% 0% 6%



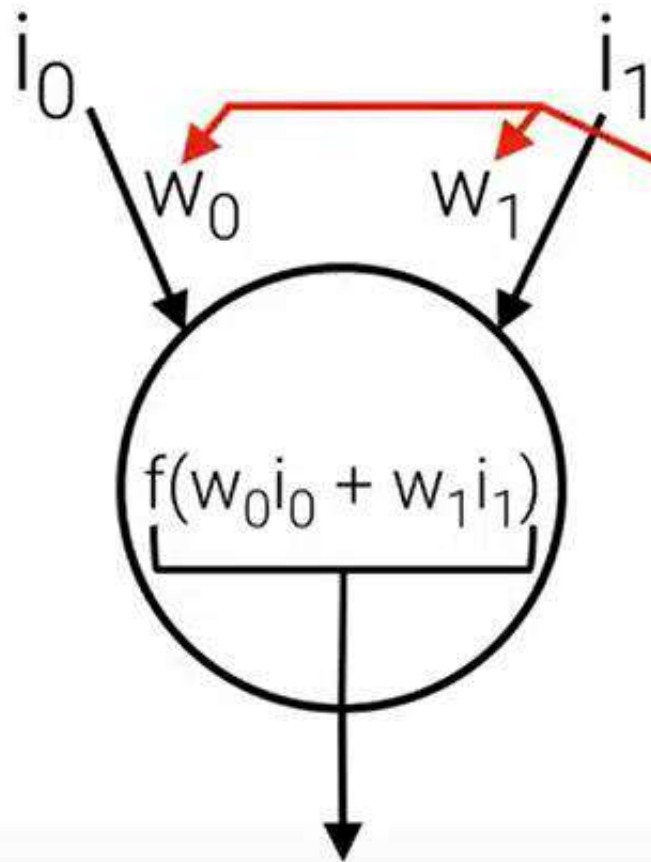




## Hyperbolic tangent (tanh)

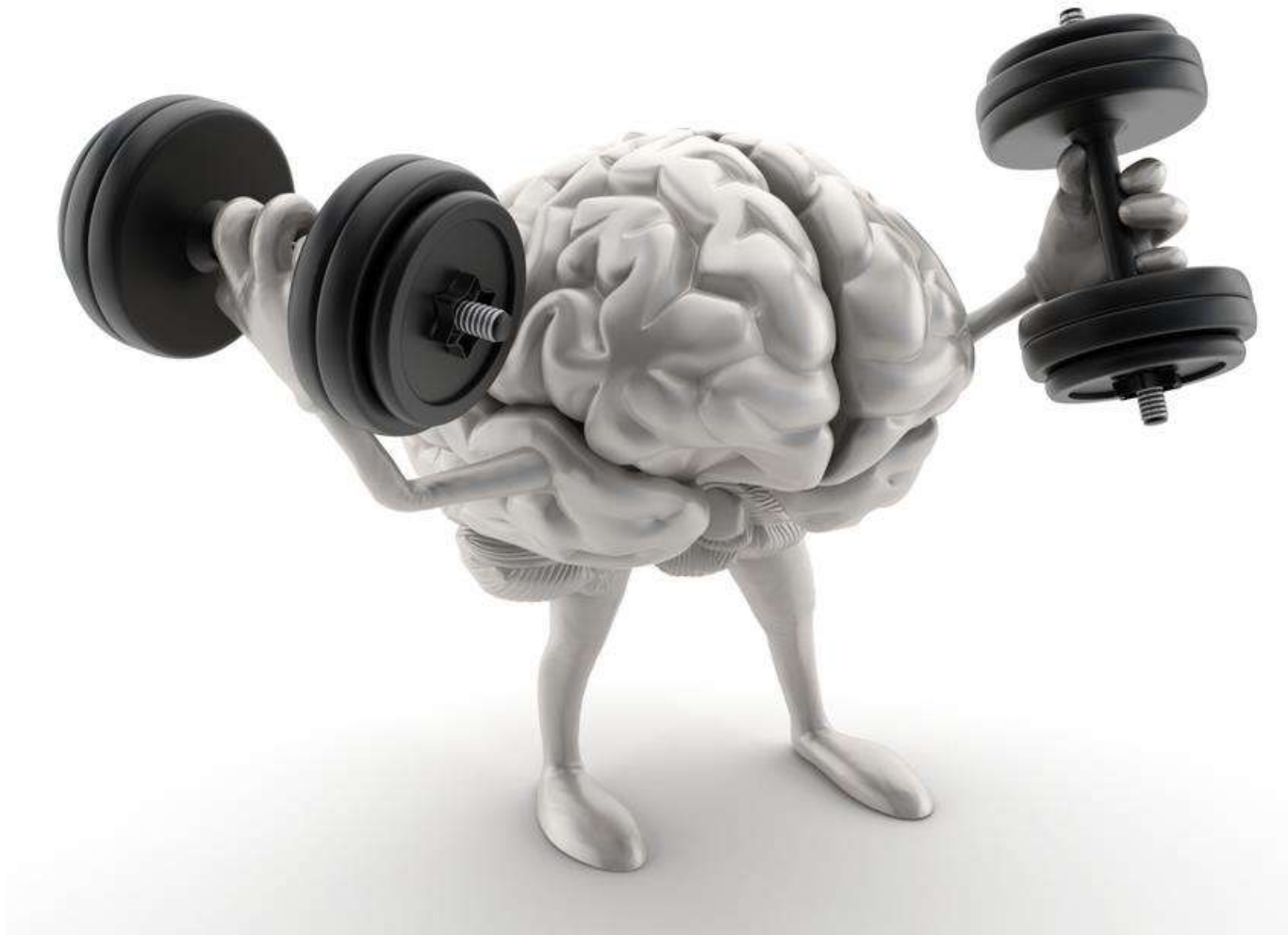


Non-linear functions allow us to capture non-linear relationships between inputs and outputs!

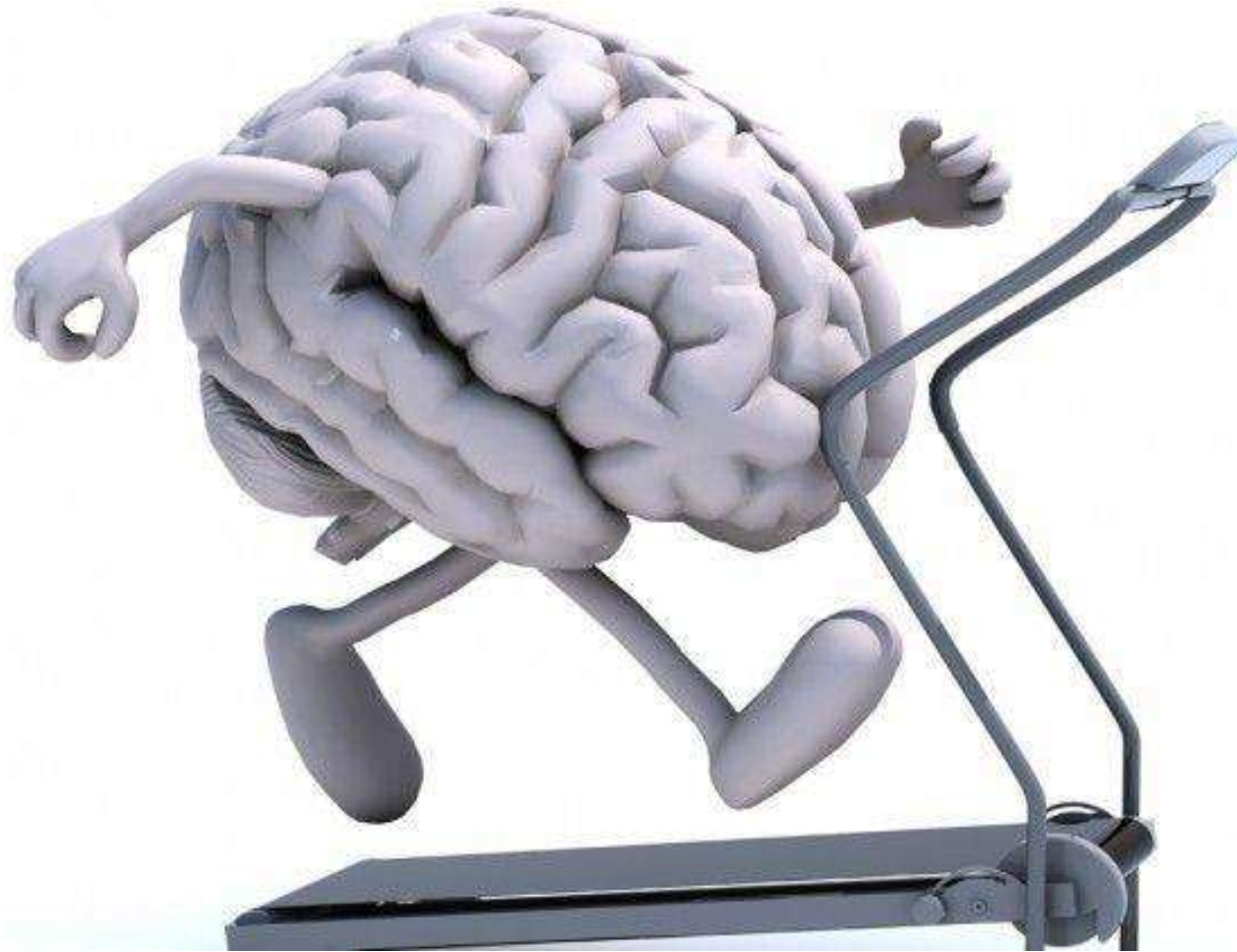


**How do we pick  
these weights?**

## Treinamento por condicionamento



## Treinamento por condicionamento





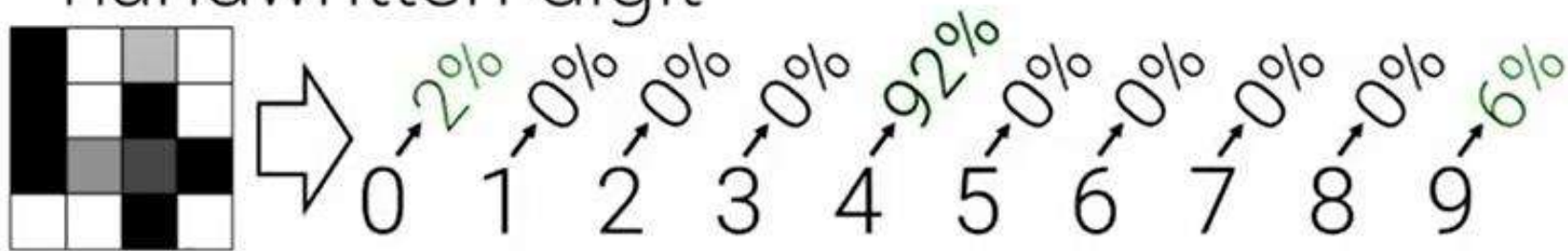
# Gradient Descent

- **Goal:** Build a classifier to identify a handwritten digit
- **Input:** A vector representing the pixels in an image
- **Output:** A vector of ten probabilities (one for each possible digit)

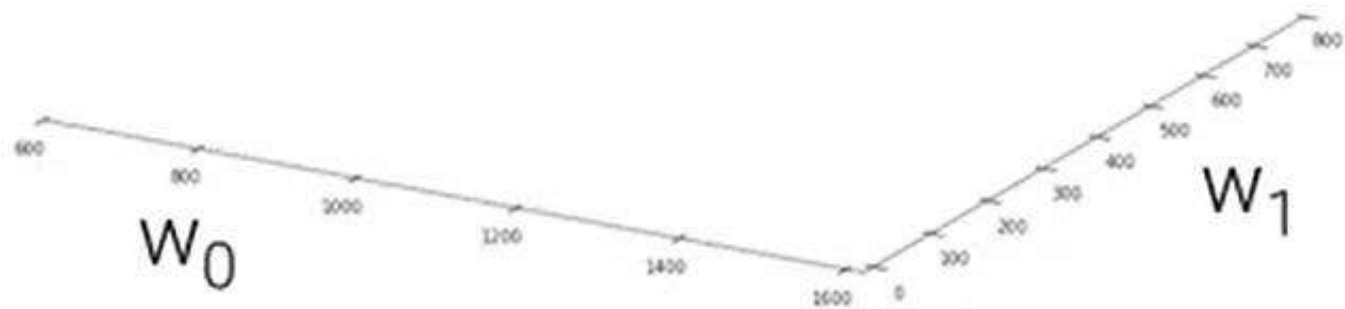


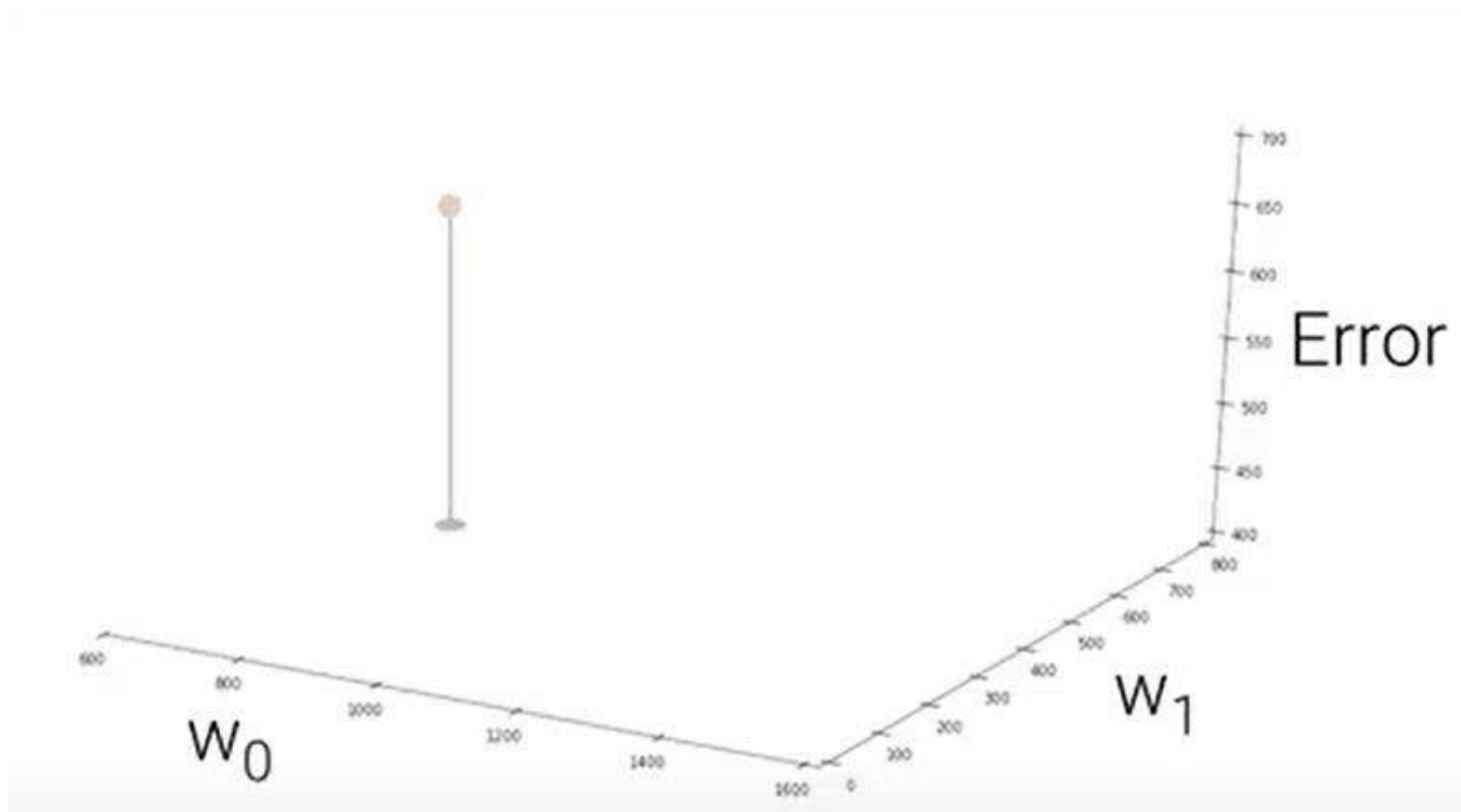
# Gradient Descent

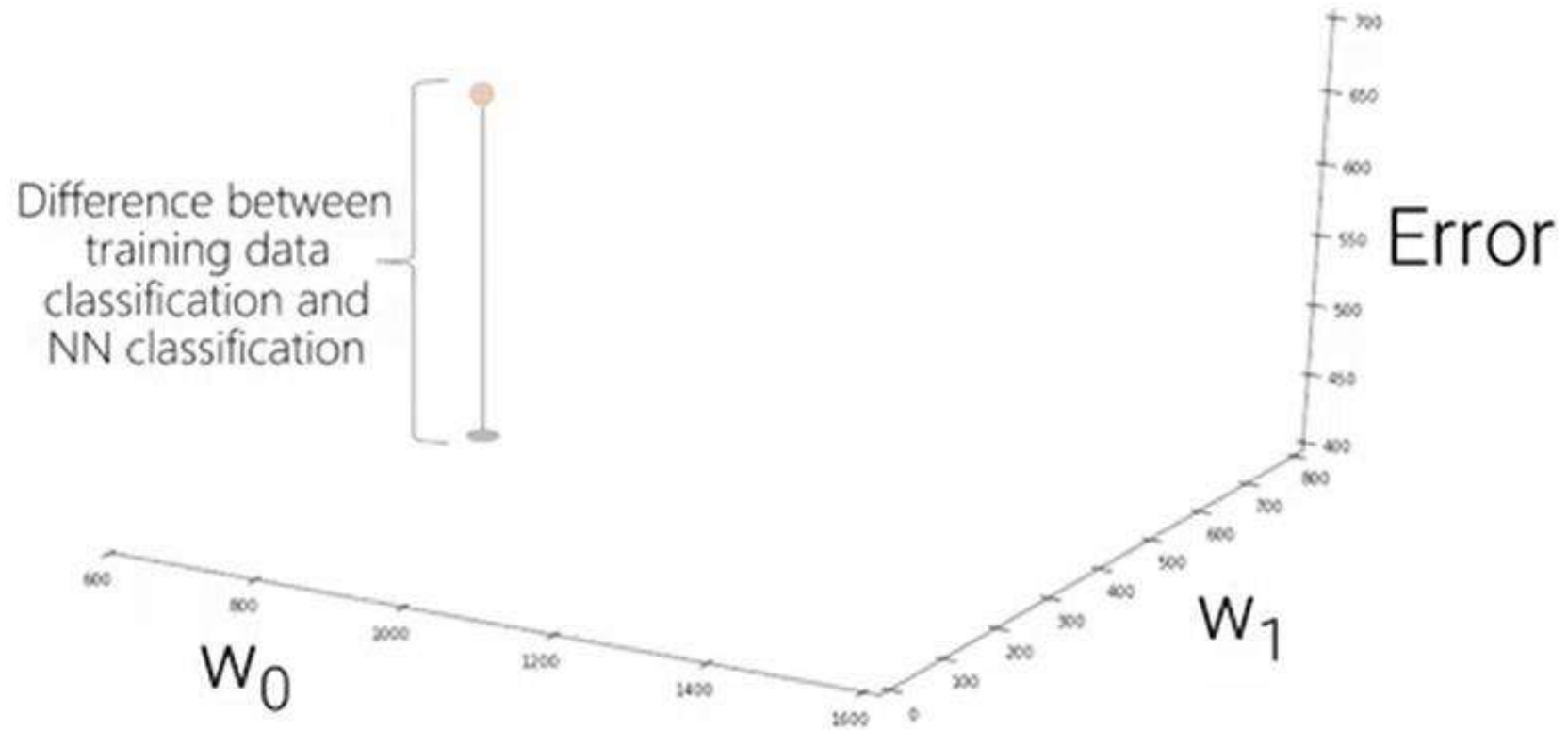
- Goal: Build a classifier to identify a handwritten digit

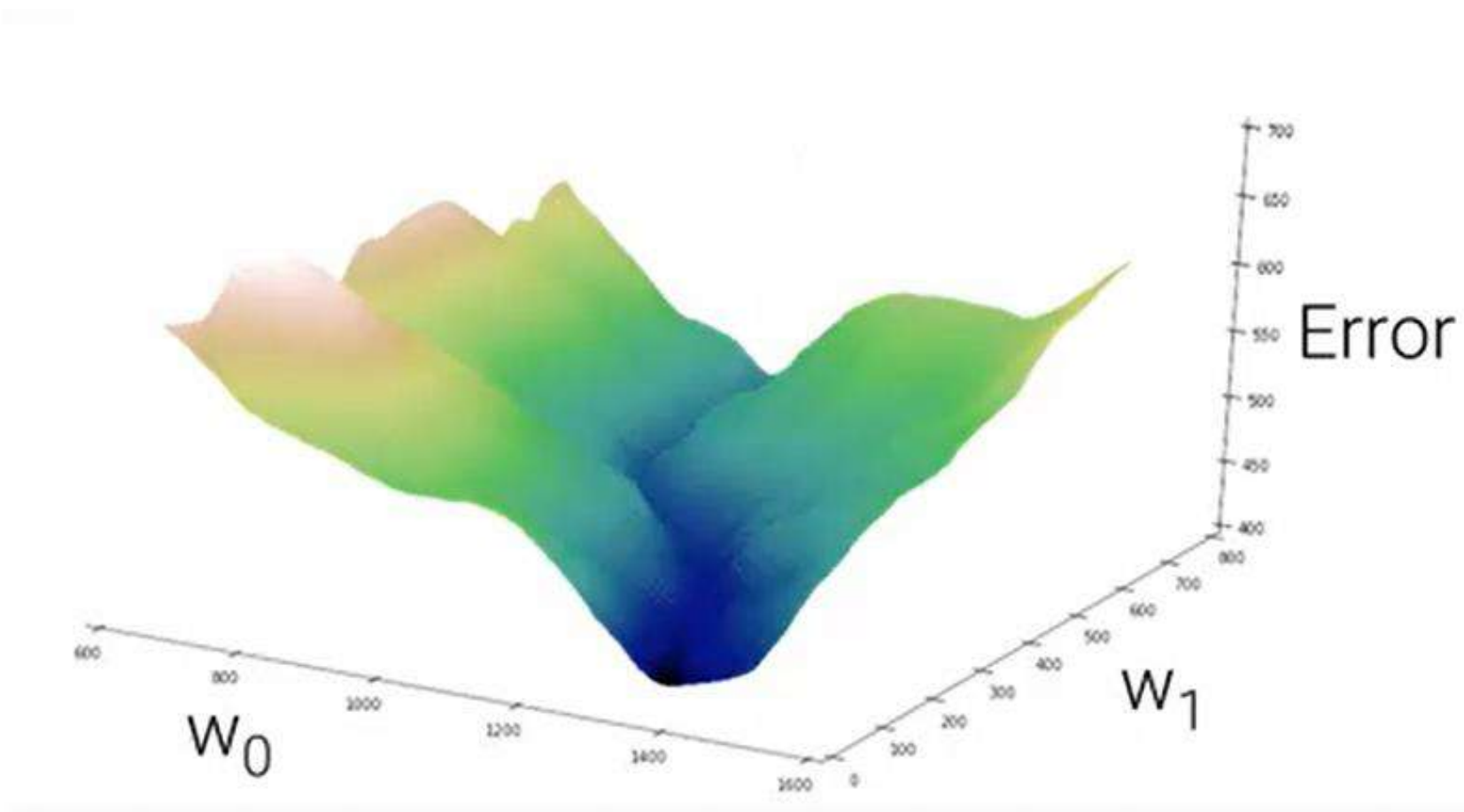


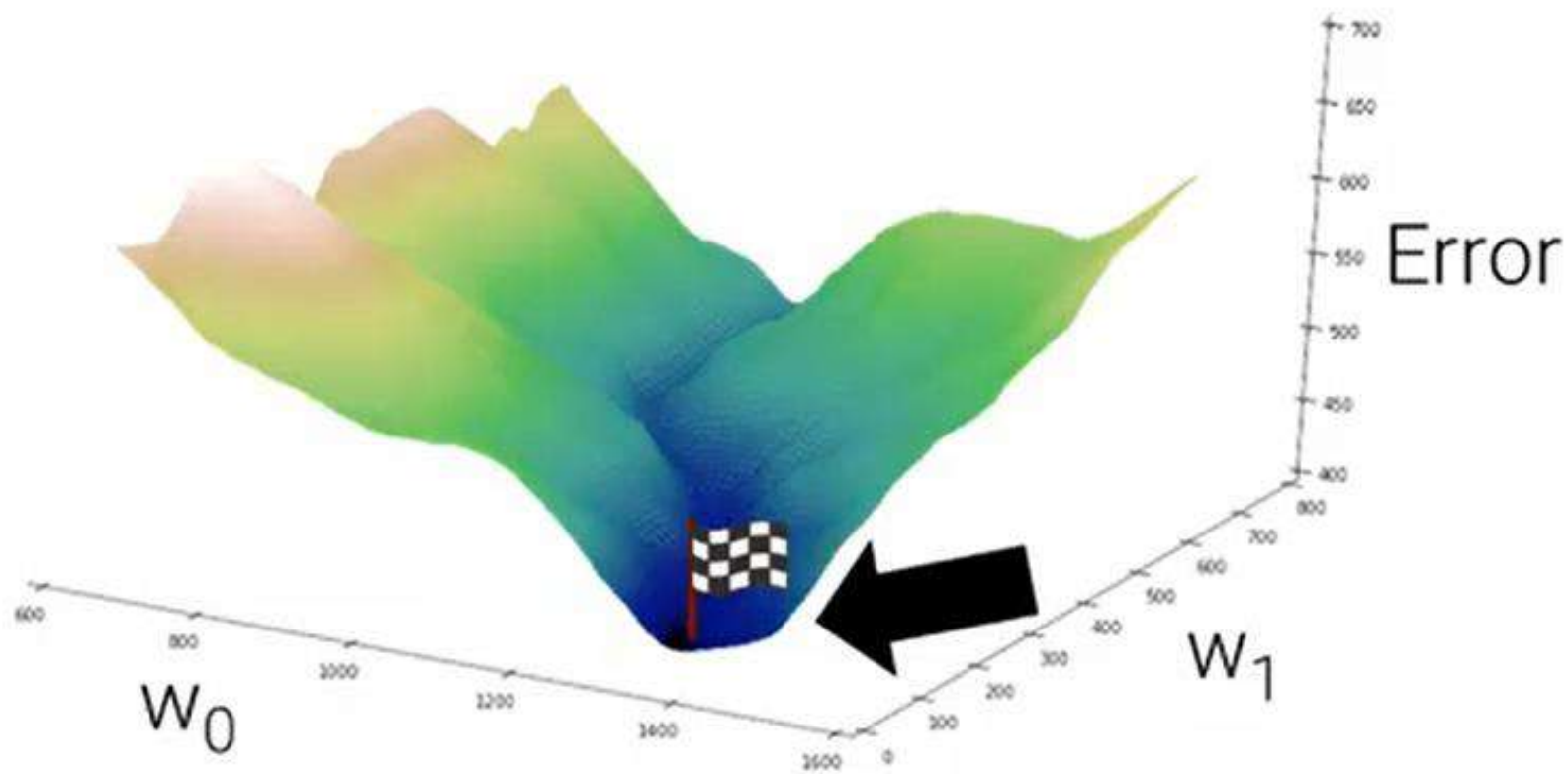
- Output: A vector of ten probabilities (one for each possible digit)

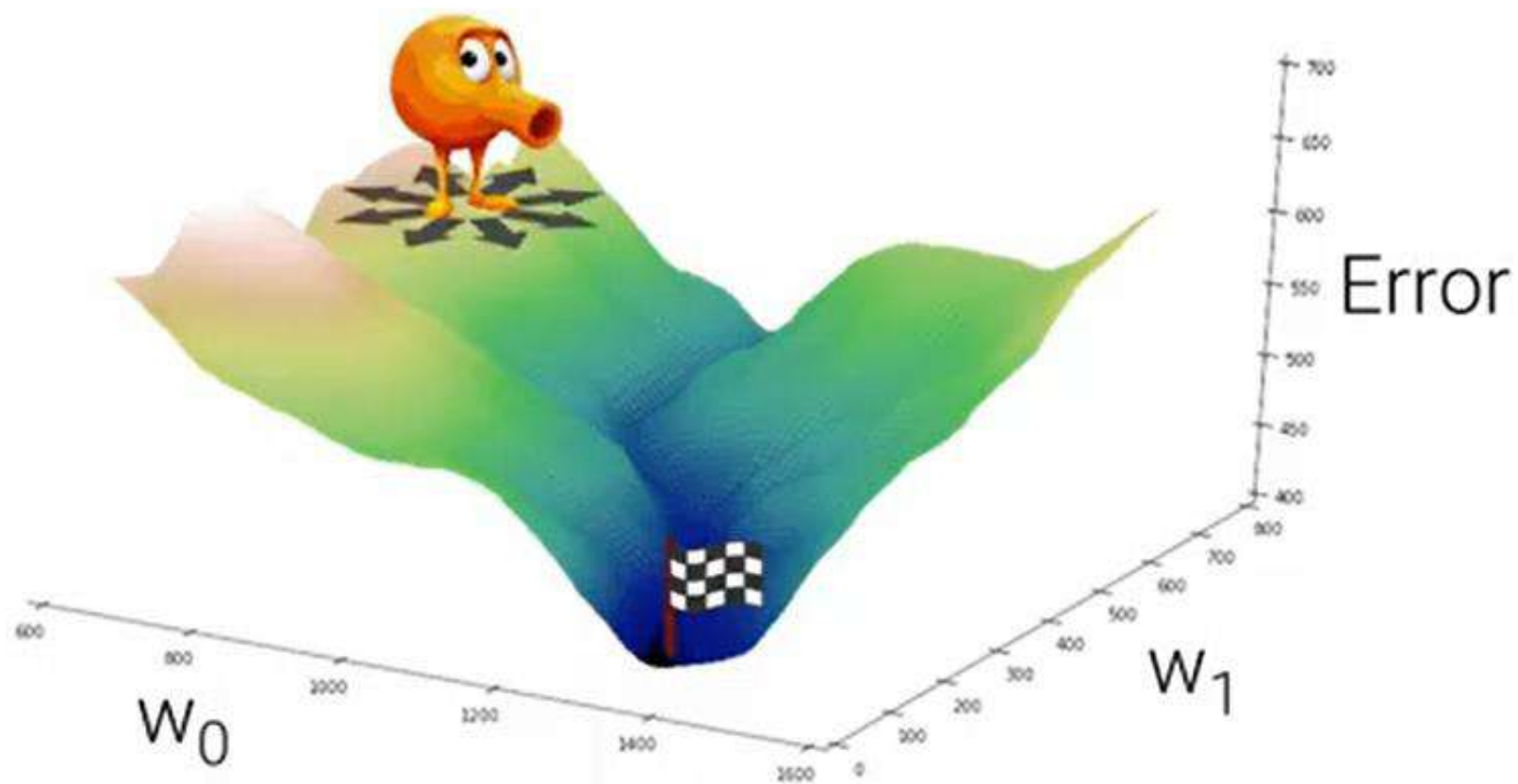


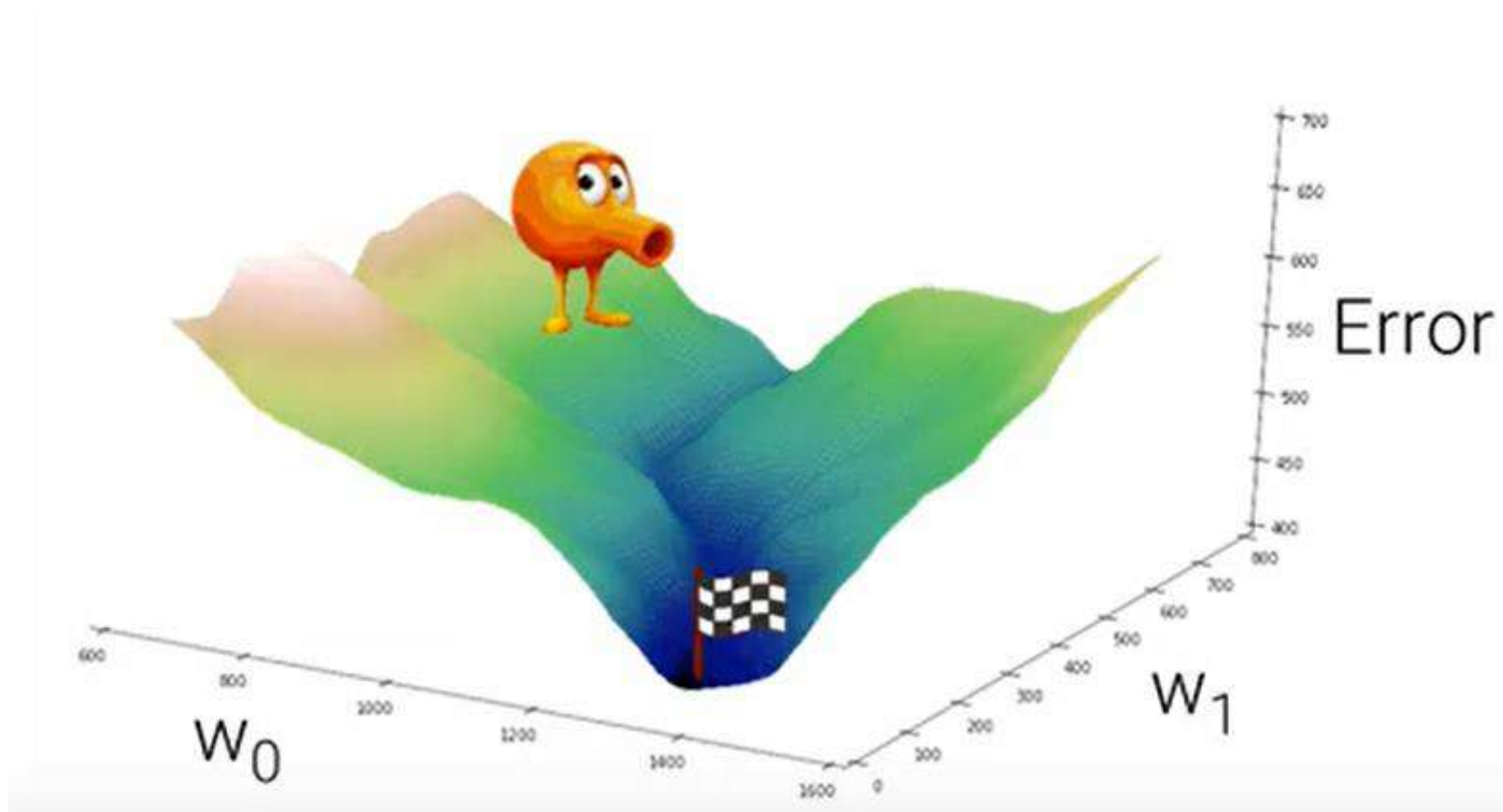




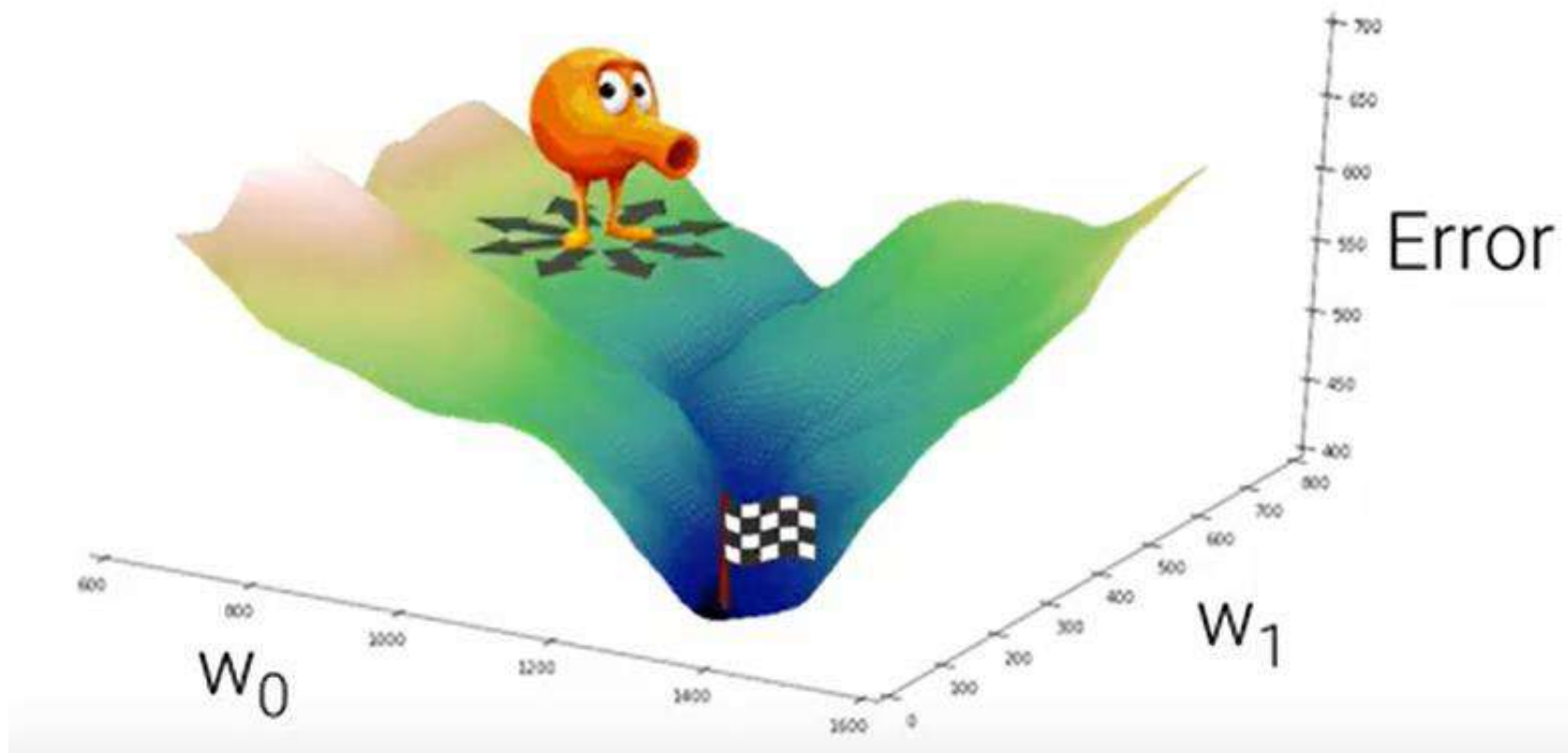


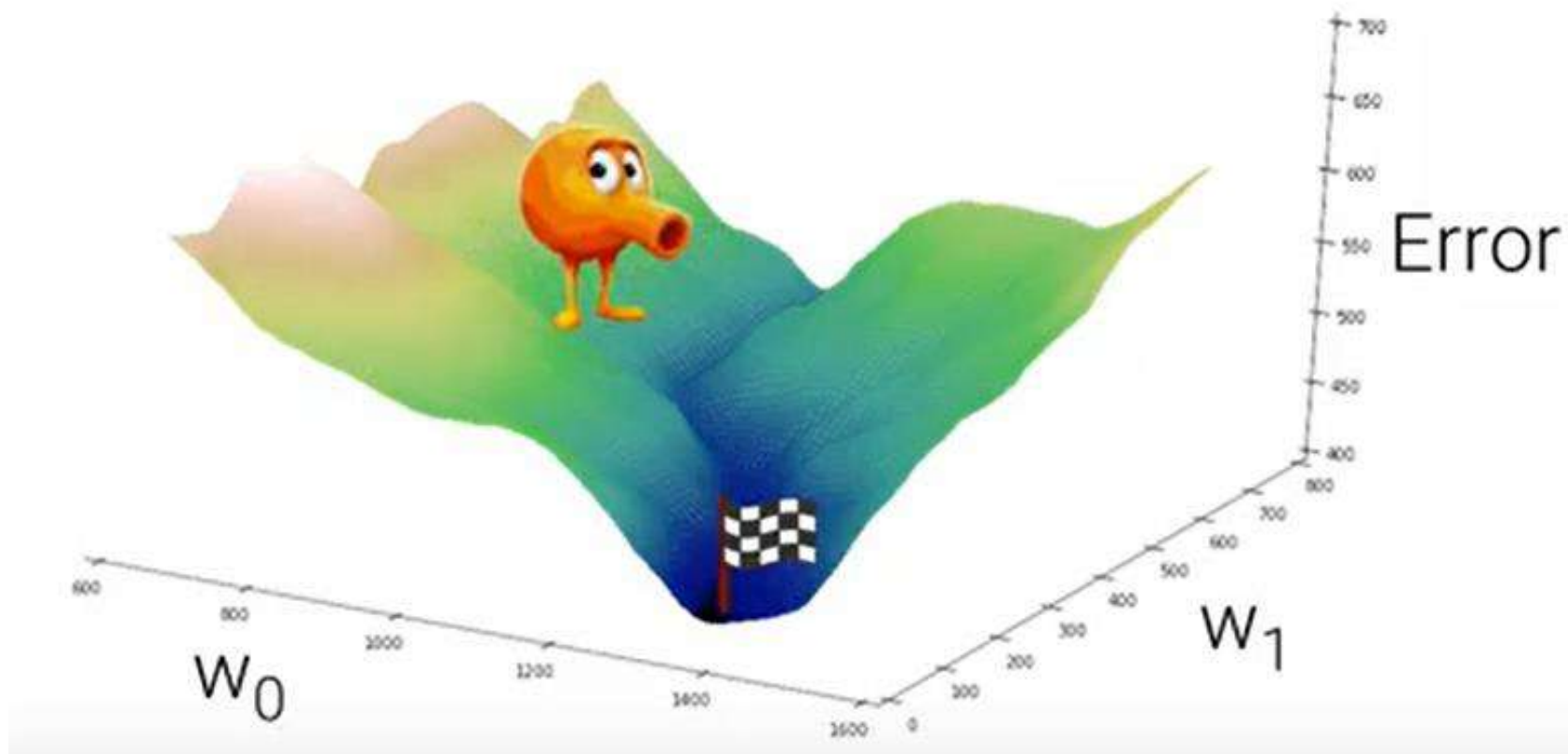


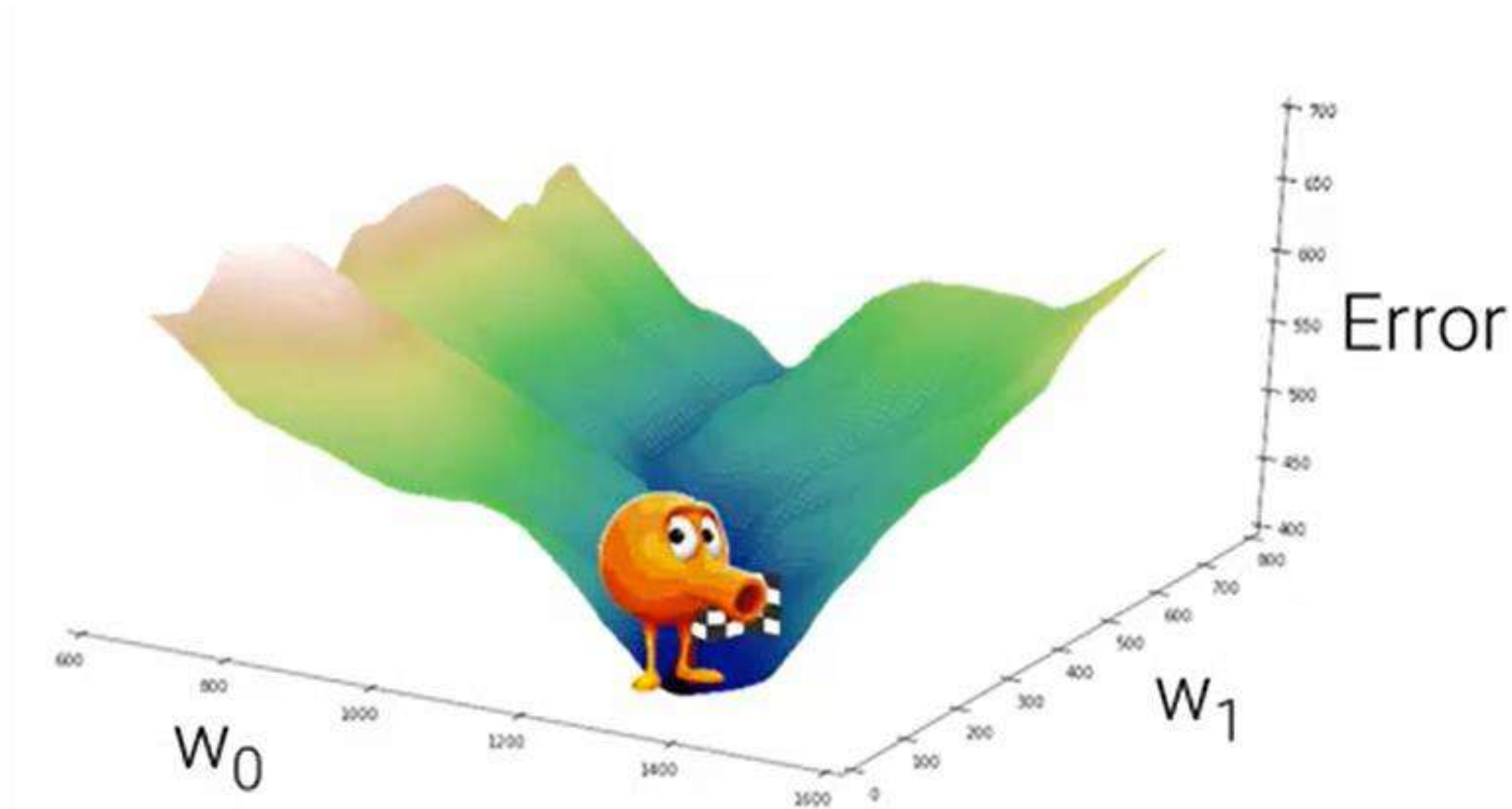






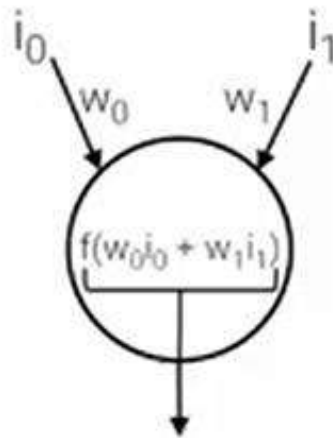




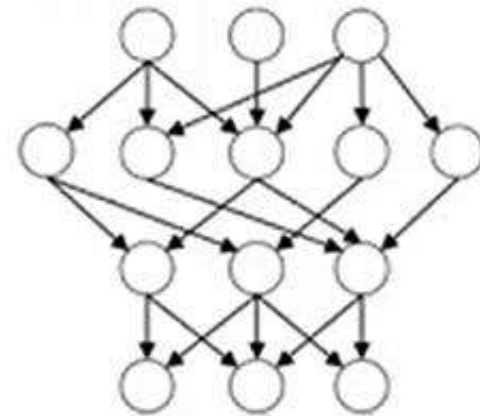


## A questão dos hiperparâmetros

What's the best  $f()$  to use?



How many layers should we use?



What should the size of a descent step be?



How big should each layer be?



Figura 5 – Exemplo de amostras da base de dados MNIST, com resolução  $28 \times 28$ .

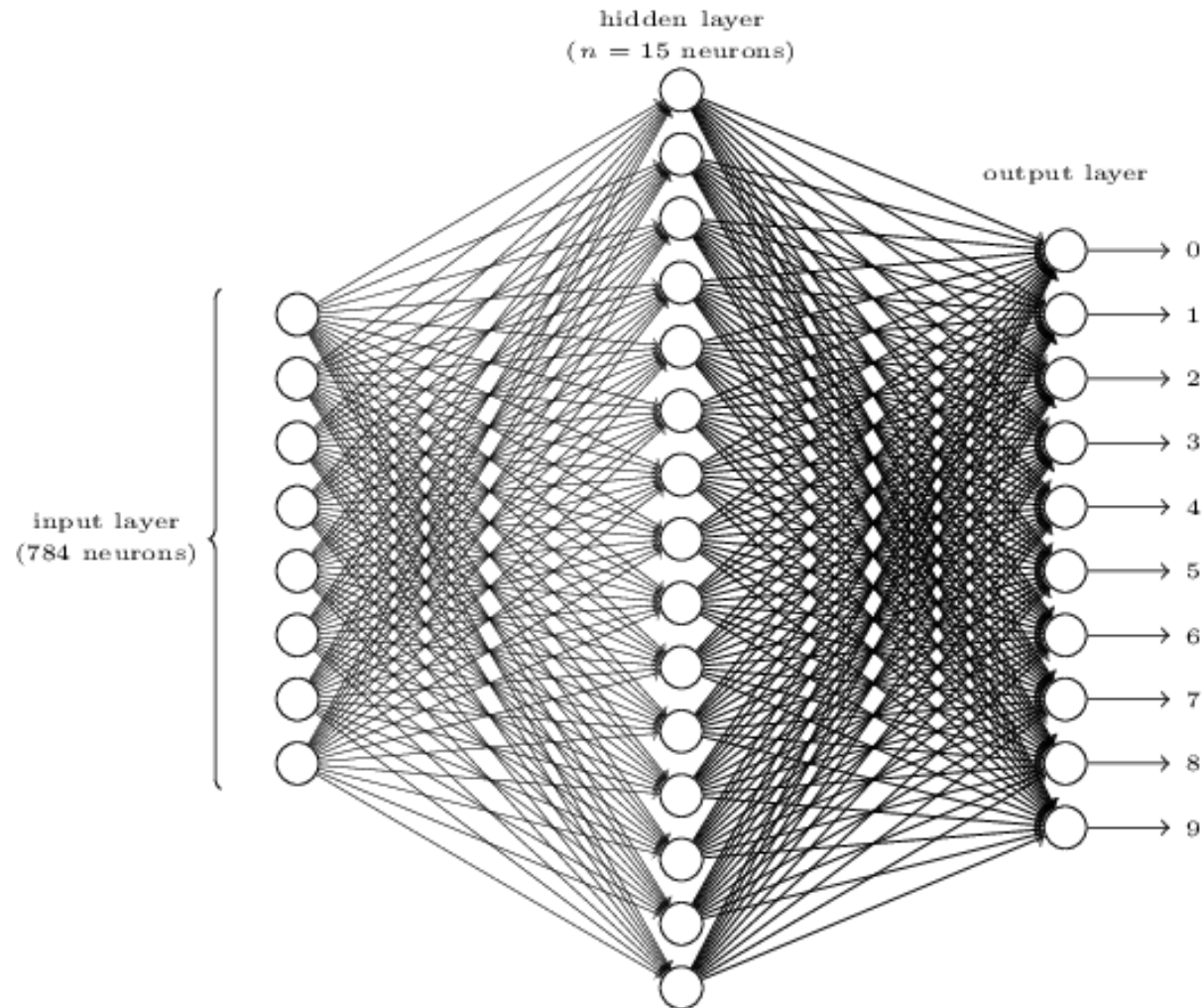


Figura 6 – Rede neural MLP com uma camada intermediária [Extraído de Nielsen, M.A. “Neural Networks and Deep Learning”, Determination Press, 2015].

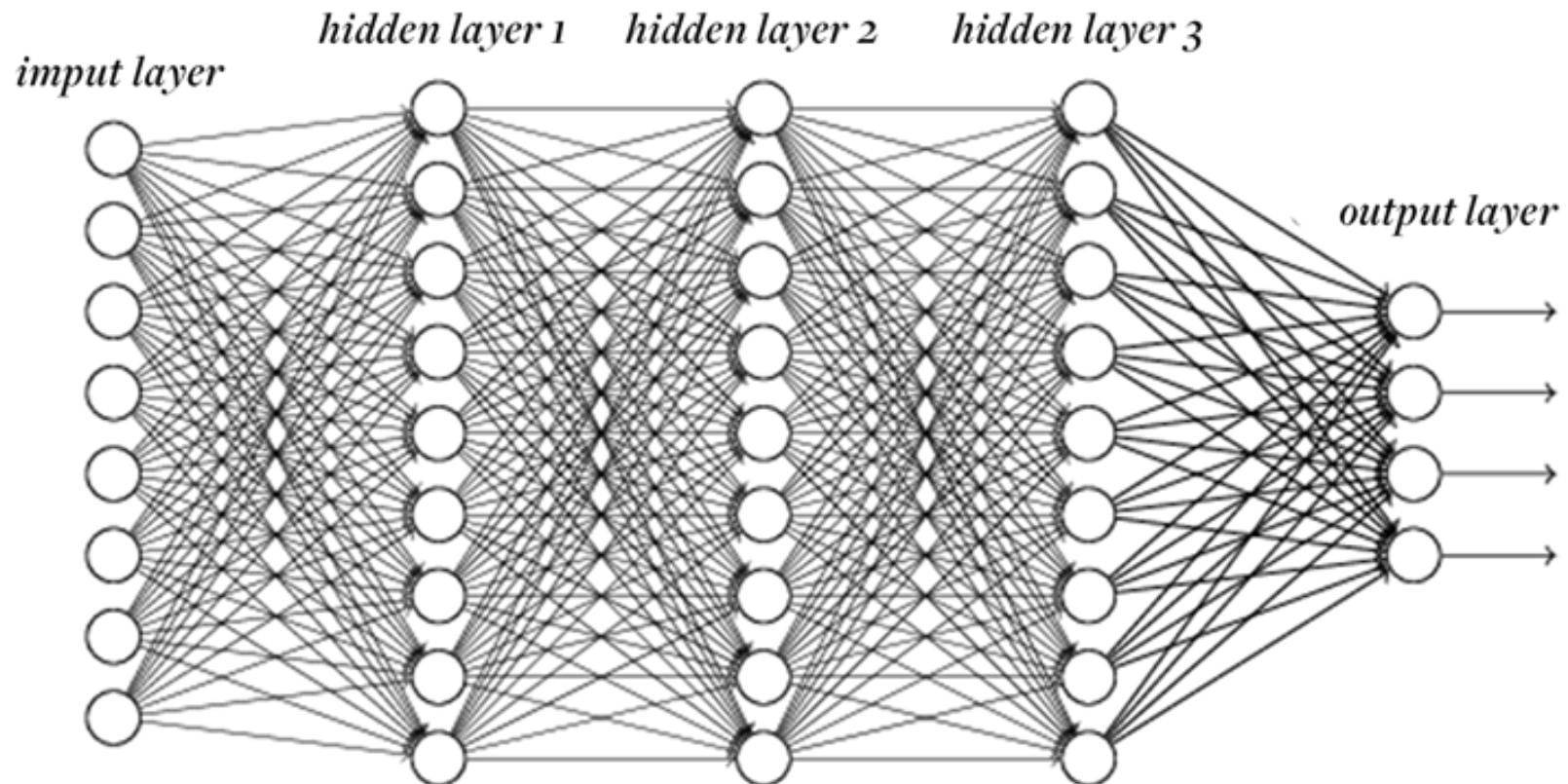


Figura 7 – Rede neural MLP com três camadas intermediárias [Extraído de <http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable> ].



## 6 A inteligência artificial já é uma realidade

- “Artificial Intelligence is the new electricity.”

Andrew Ng

- Veículos autônomos estão virando a esquina de nossas ruas, máquinas fazem tradução automática de texto, rotulam imagens de forma automática e transcrevem mensagens de voz.
- Empresas bilionárias são sustentadas por técnicas de inteligência artificial.



Tarefa: Compare a capacidade intelectual de [um ser humano] com aquela de [um ser humano + um smartphone].



## Cérebros de pessoas sociáveis têm áreas mais desenvolvidas, diz estudo

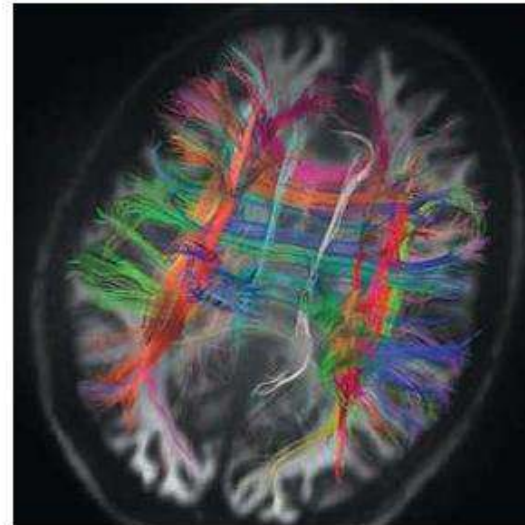
Do UOL, em São Paulo 23/11/2013 05h00

Pessoas que cultivam amizades e fazem novos amigos com facilidade apresentam cérebros com estruturas diferenciadas, aponta estudo conduzido pela universidade canadense McGill e divulgado durante conferência da Sociedade de Neurociência, nos Estados Unidos.

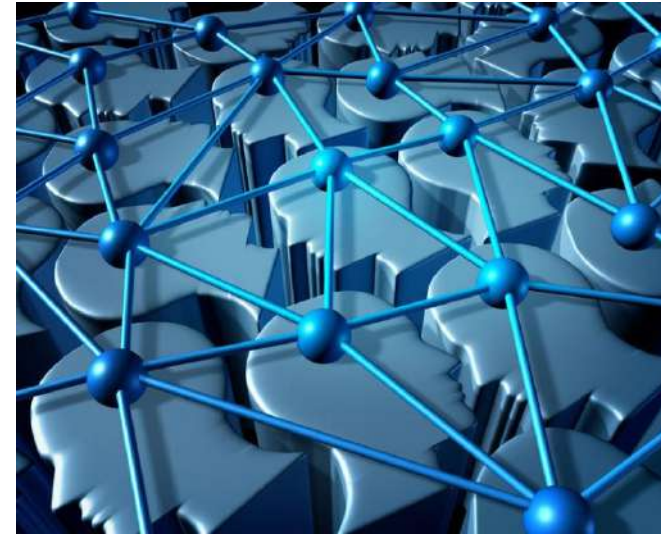
Segundo a pesquisa, áreas do cérebro de pessoas notavelmente sociáveis exibem dimensões maiores e conexões mais marcantes do que cérebros de pessoas introspectivas.

Para chegar à descoberta, os idealizadores do estudo analisaram imagens dos cérebros de 18 participantes. Antes de terem seus cérebros escaneados, os indivíduos analisados eram questionados sobre quantas interações sociais haviam tido nos último mês —isso ajudava os pesquisadores a estabelecer o tamanho de suas redes de contato.

Center for Biomedical Imaging, Massachusetts General Hospital



Os caminhos de fibra de um cérebro humano mostram como as principais vias dos lobos frontais se interligam, organizando-se em ângulos retos.



- A figura da esquerda não é realista, mas na prática é isso que o smartphone consegue fazer: conectar cérebros. A figura da direita remete ao conceito de super-inteligência, formada pela conectividade de múltiplos cérebros. Trata-se de um requisito fundamental para lidar com a expansão tecnológica.
- Além disso, não estamos apenas conectando cérebros com cérebros, mas cérebros com máquinas (coisas) e também máquinas com máquinas. O próximo passo, já em plena implementação, é fazer essas máquinas aprenderem com a linguagem natural humana e a usarem para se comunicar com humanos.



Aprendizado de máquina está fazendo pela inteligência artificial o que a alquimia fez pela química.

- De fato, já é um grande desafio nominar um ramo de atuação científica que ainda não tenha se beneficiado de soluções de aprendizado de máquina. É por isso que alguns analistas já apontam o aprendizado de máquina como uma nova forma de alfabetização.

## 7 Computação Digital e Neurocomputação

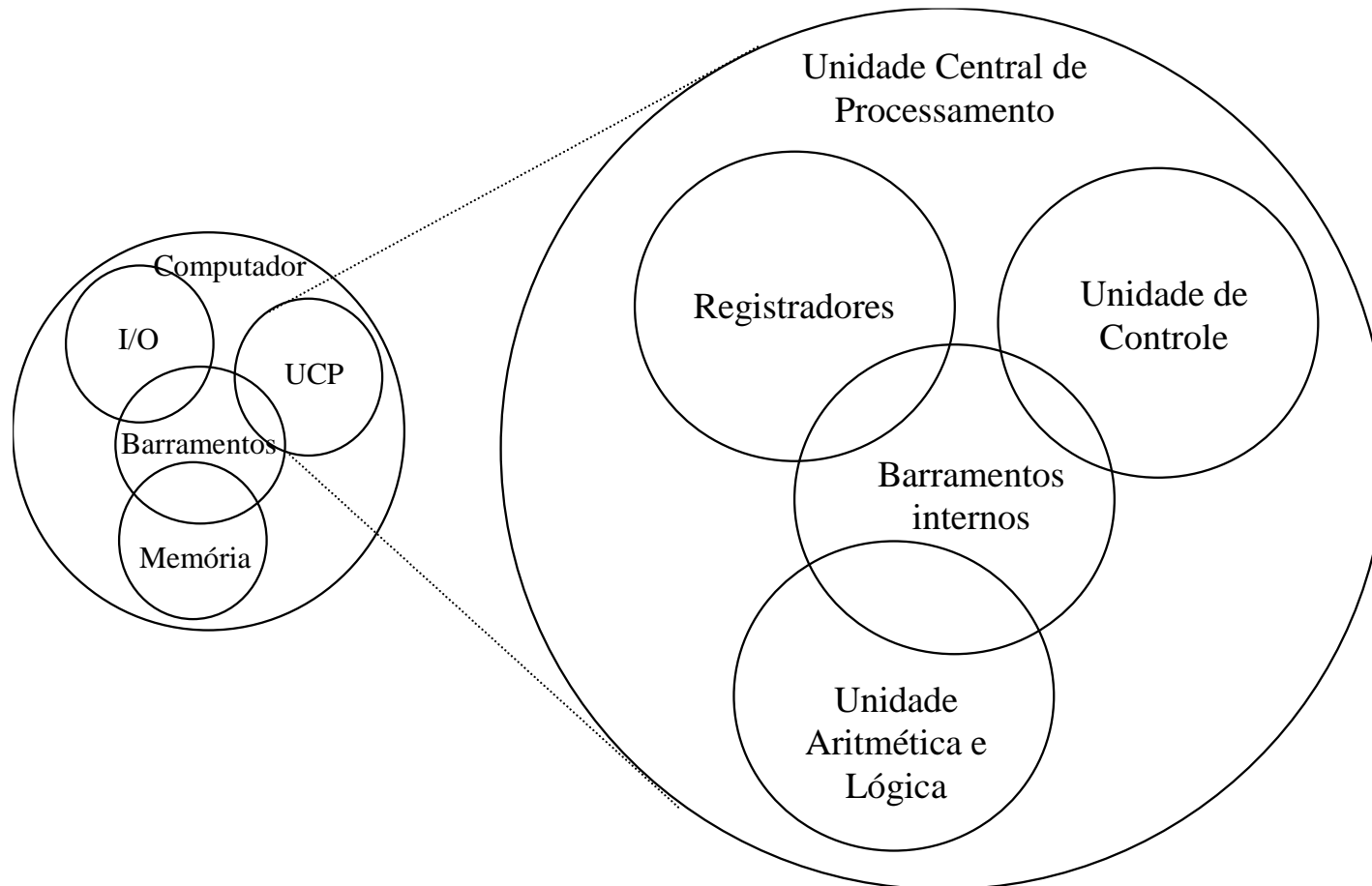


Figura 8 – Princípio da computação digital

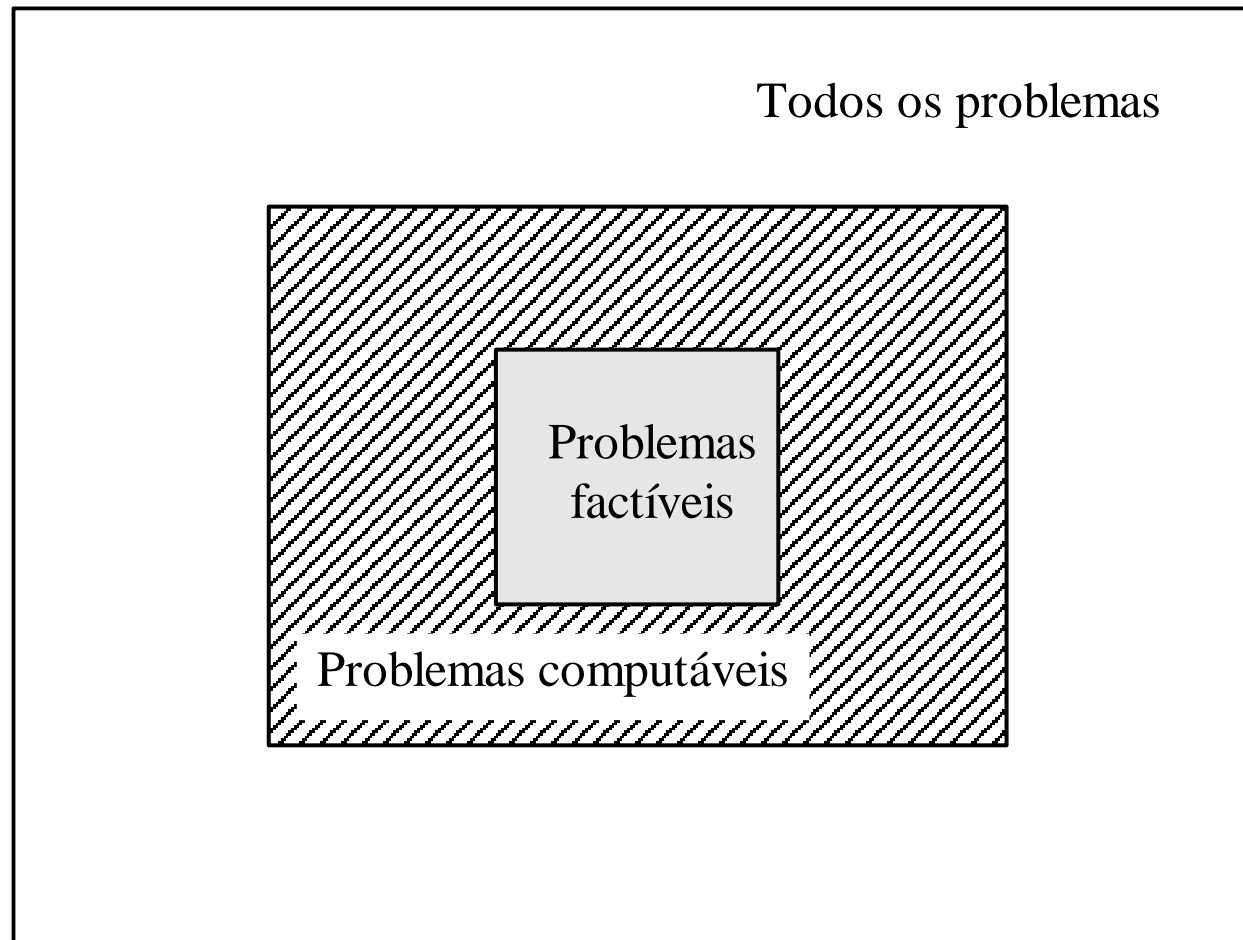


Figura 9 – Complexidade computacional (foco na computação digital)

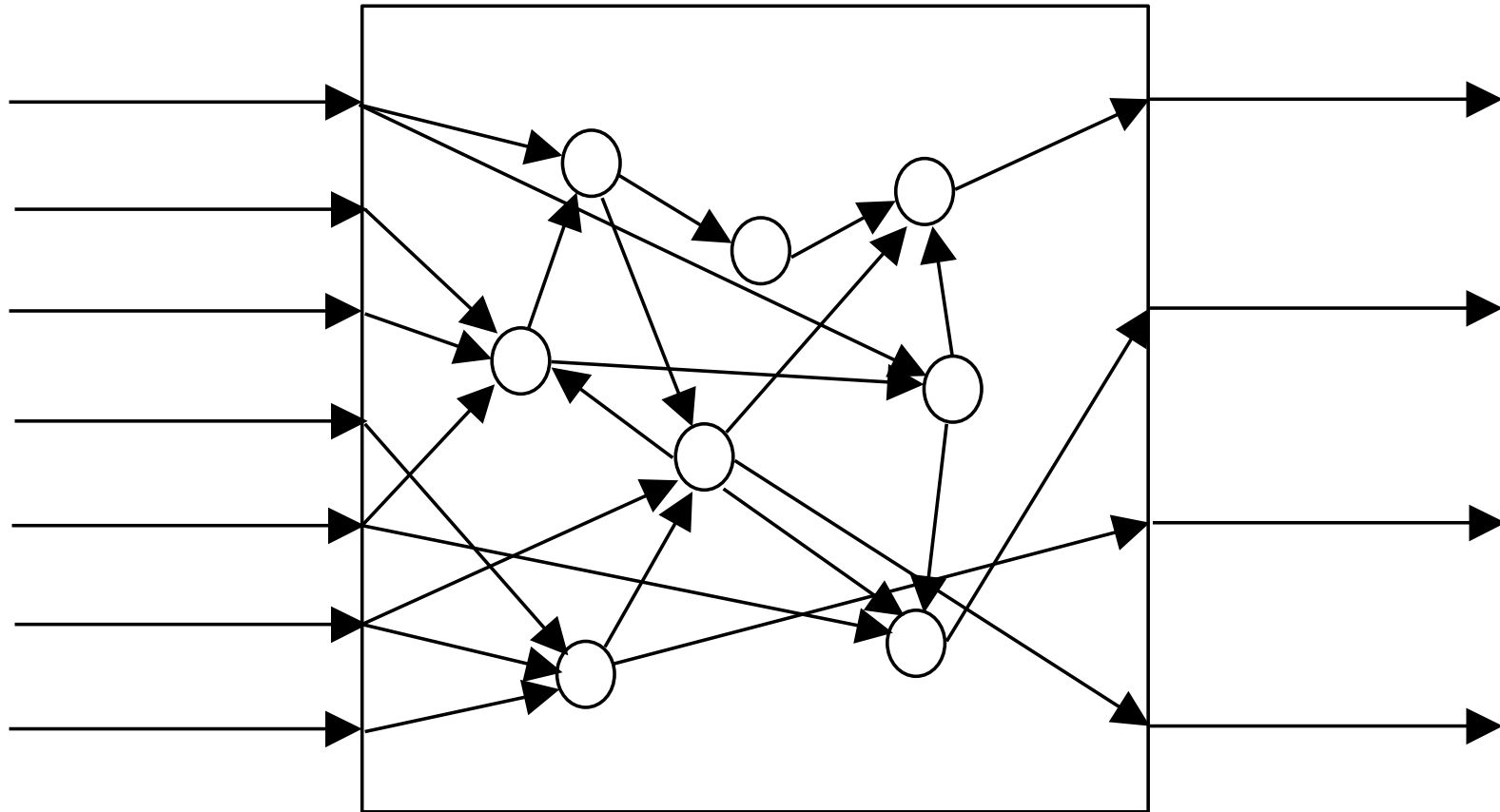


Figura 10 – Exemplo genérico de um neurocomputador

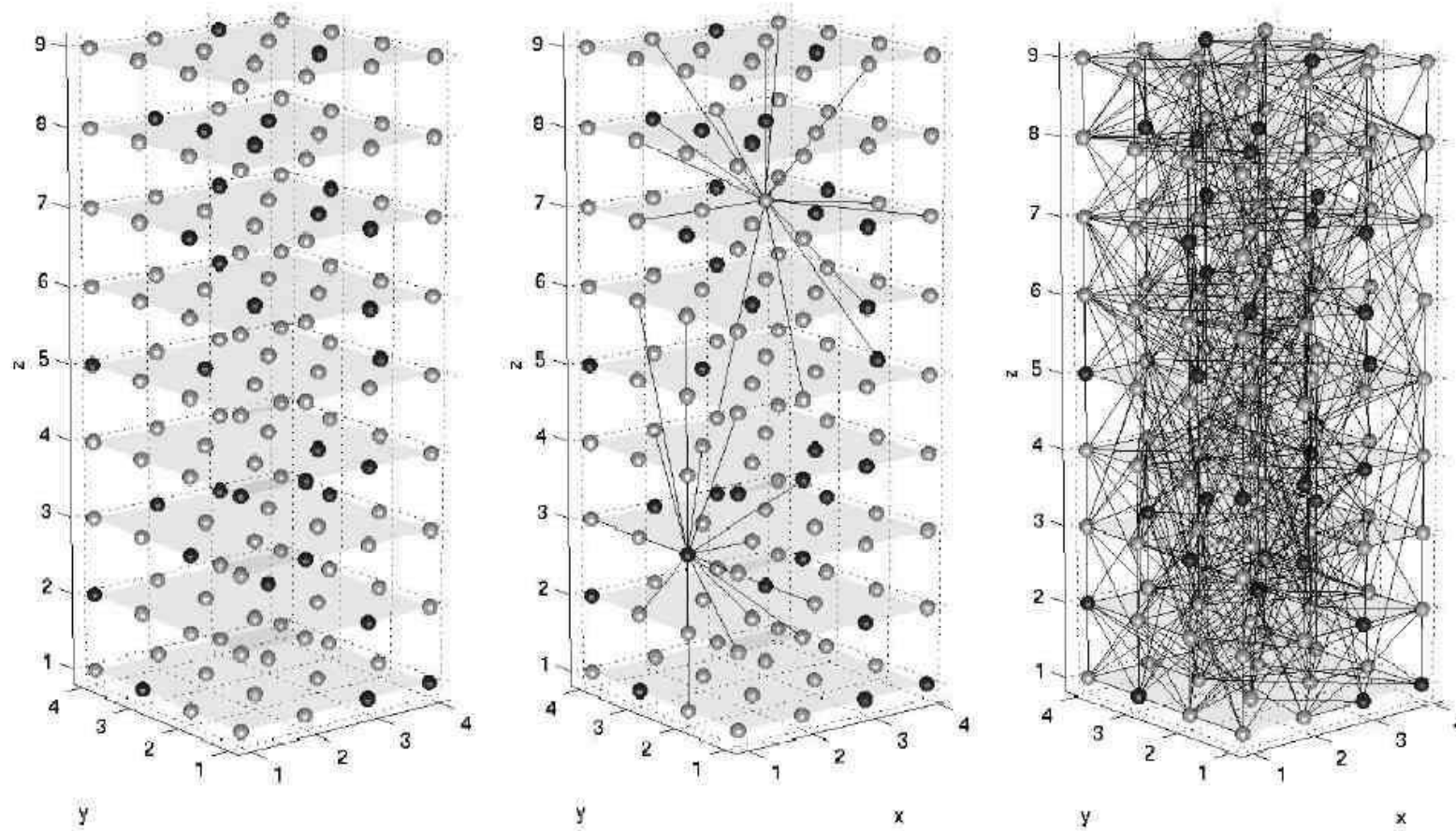


Figura 11 – Paradigma conexionista: Rede = nós + conexões



## **8 Lei de Moore e a expansão tecnológica**

- Qual é a porcentagem das pessoas que já “passaram pela face da Terra” e que ainda estão vivas?
  
- Qual é a porcentagem de pessoas que se tornaram cientistas, em toda a história da humanidade, e que ainda estão vivas?
  
- A resposta a essas duas questões ajuda a entender melhor o mundo em que vivemos. Esse segundo número foi apresentado em [David Goodstein, “The Big Crunch” Proceedings of the 48th NCAR Symposium, Portland, 1994] e se mantém por pelo menos 300 anos.



- Qual é a porcentagem das pessoas que já “passaram pela face da Terra” e que ainda estão vivas?

6% a 9%

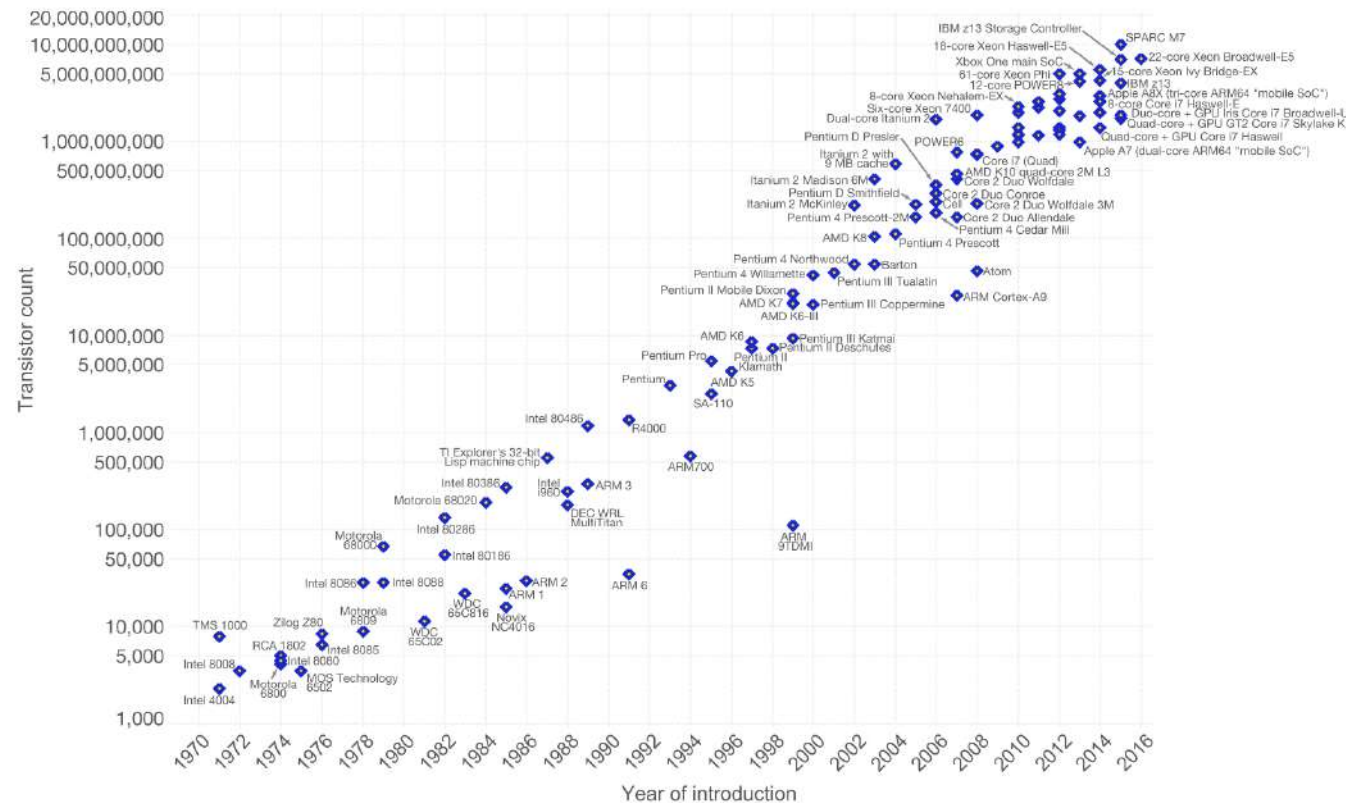
- Qual é a porcentagem de pessoas que se tornaram cientistas, em toda a história da humanidade, e que ainda estão vivas?

~92%

- A resposta a essas duas questões ajuda a entender melhor o mundo em que vivemos. Esse segundo número foi apresentado em [David Goodstein, “The Big Crunch” Proceedings of the 48th NCAR Symposium, Portland, 1994] e se mantém por pelo menos 300 anos.

## Moore's Law – The number of transistors on integrated circuit chips (1971-2016) Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.



Data source: Wikipedia ([https://en.wikipedia.org/wiki/Transistor\\_count](https://en.wikipedia.org/wiki/Transistor_count))

The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

Figura 12 – Lei de Moore vigora por 5 décadas



Figura 13 – O efeito da miniaturização dos transistores

- A lei de Moore não é uma lei do mundo físico, como muitos talvez pensem. Ela é uma conquista tecnológica e uma indicação de como funciona o mundo corporativo.

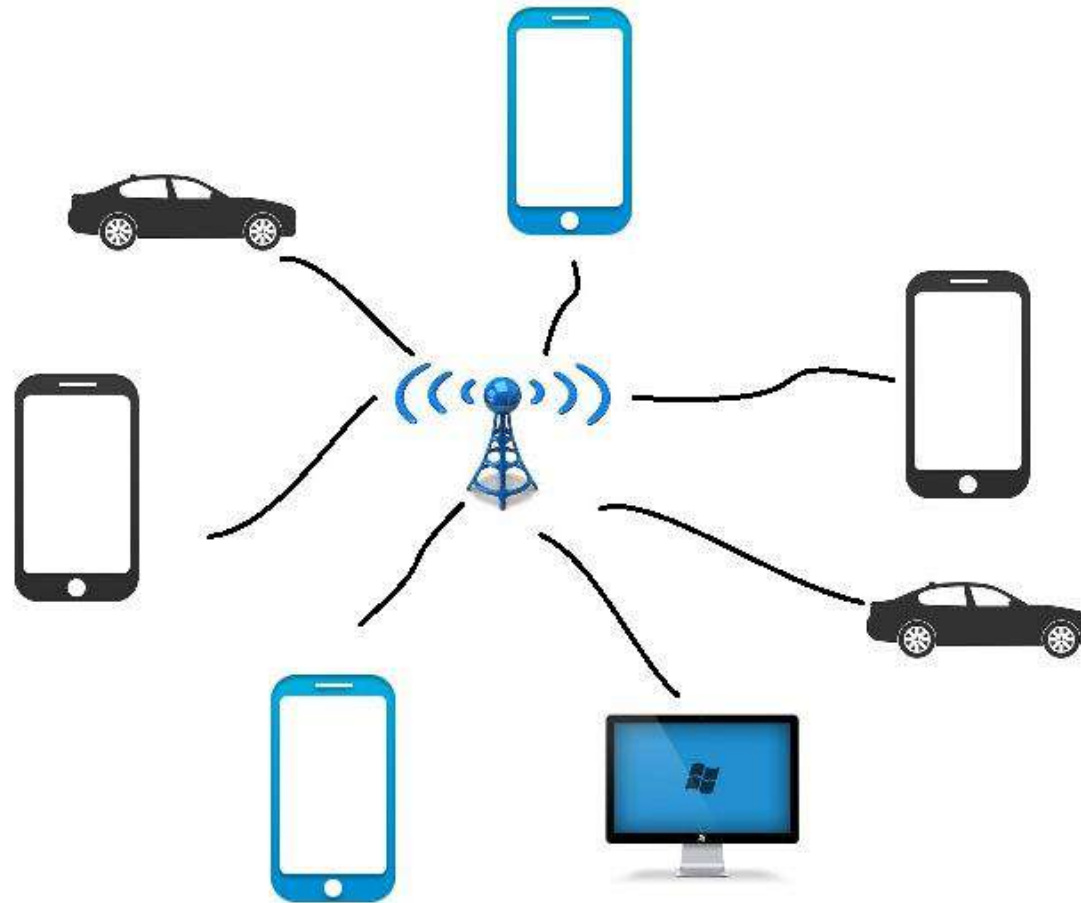


Figura 14 – Outras conquistas tecnológicas: a comunicação sem fio e os dispositivos móveis

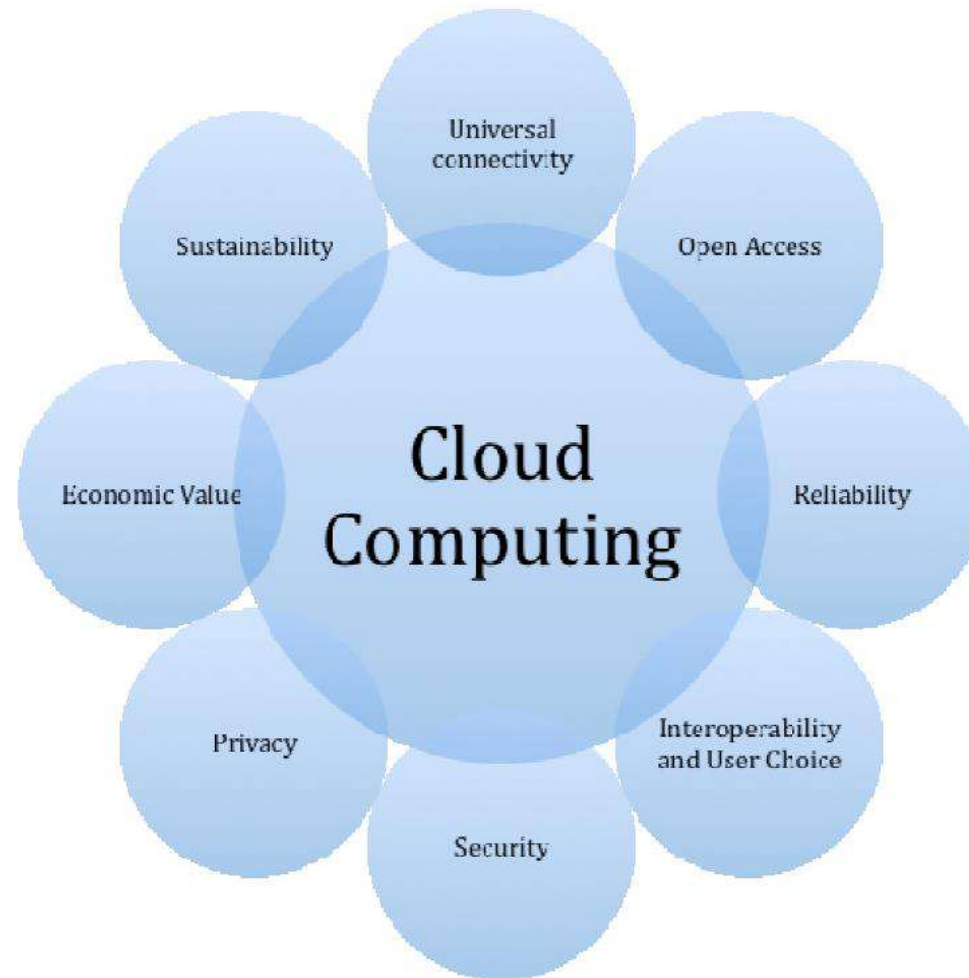


Figura 15 – Elementos fundamentais da computação em nuvem

( [https://www.researchgate.net/publication/45825704\\_The\\_Cloudy\\_Future\\_Of\\_Government\\_IT\\_Cloud\\_Computing\\_and\\_The\\_Public\\_Sector\\_Around\\_The\\_World](https://www.researchgate.net/publication/45825704_The_Cloudy_Future_Of_Government_IT_Cloud_Computing_and_The_Public_Sector_Around_The_World) )

## The Internet of Everything: Networked Connections of People, Process, Data, Things



Figura 16 – A internet de todas as coisas

( <http://honim.typepad.com/biasc/2014/06/the-internet-of-everything-course.html> )



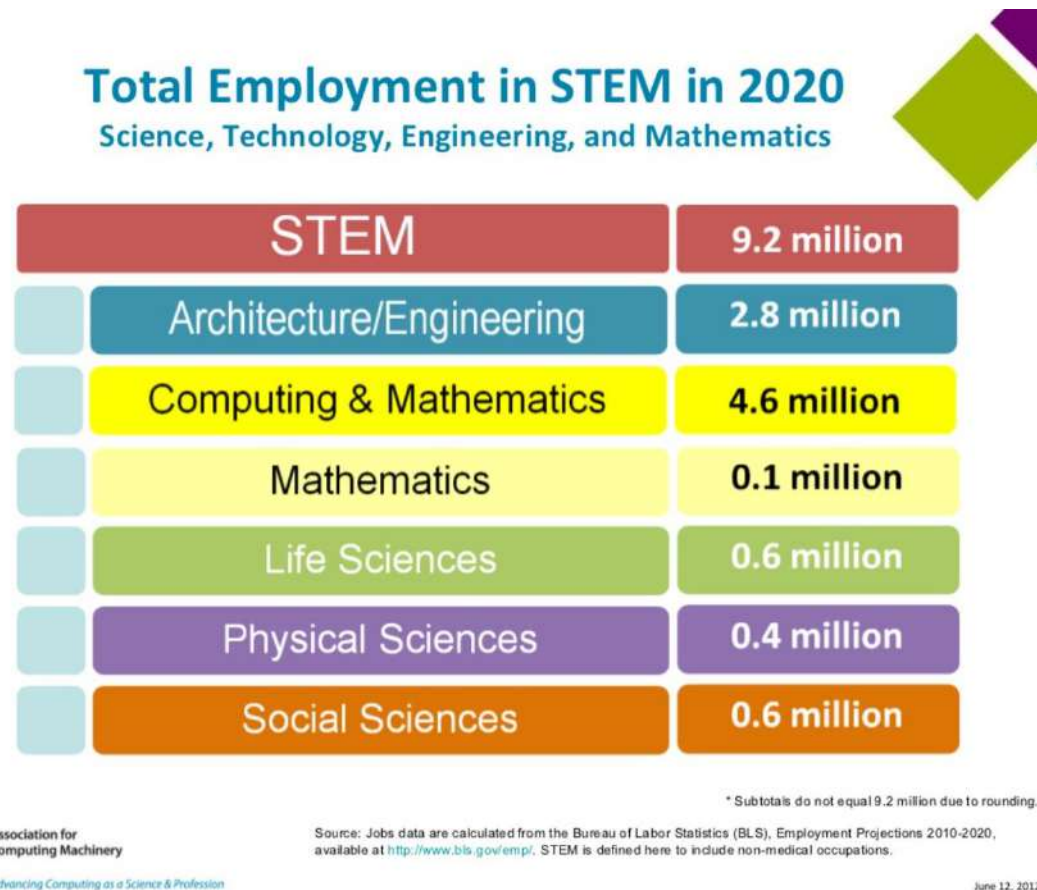


Figura 17 – Distribuição de vagas em STEM

- The U.S. Bureau of Labor Statistics describes STEM occupations as those that typically involve determining how things work and solving problems with the use of computers and scientific tools.

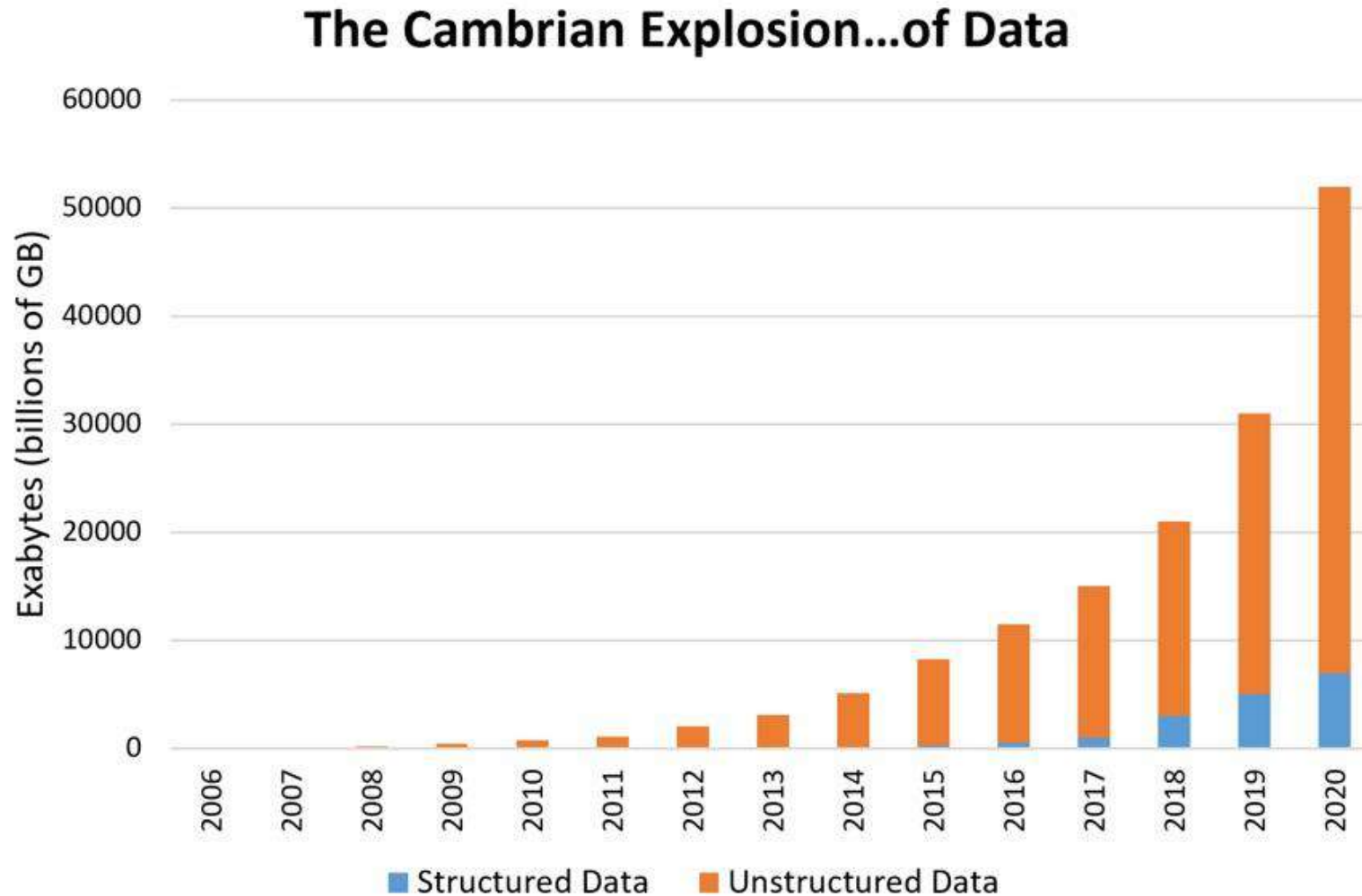


Figura 18 – Taxa de expansão na produção de dados estruturados e não-estruturados

Fonte: [https://www.eetimes.com/author.asp?section\\_id=36&doc\\_id=1330462](https://www.eetimes.com/author.asp?section_id=36&doc_id=1330462)



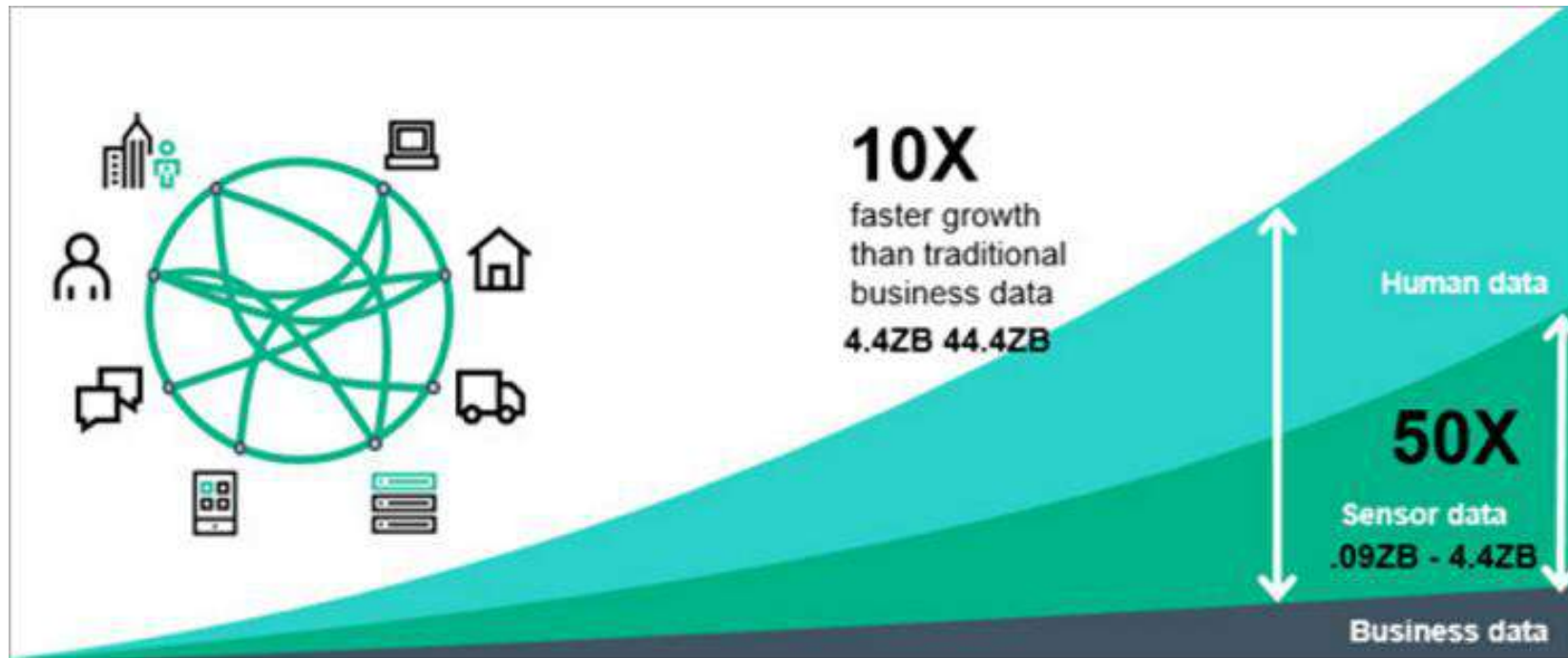
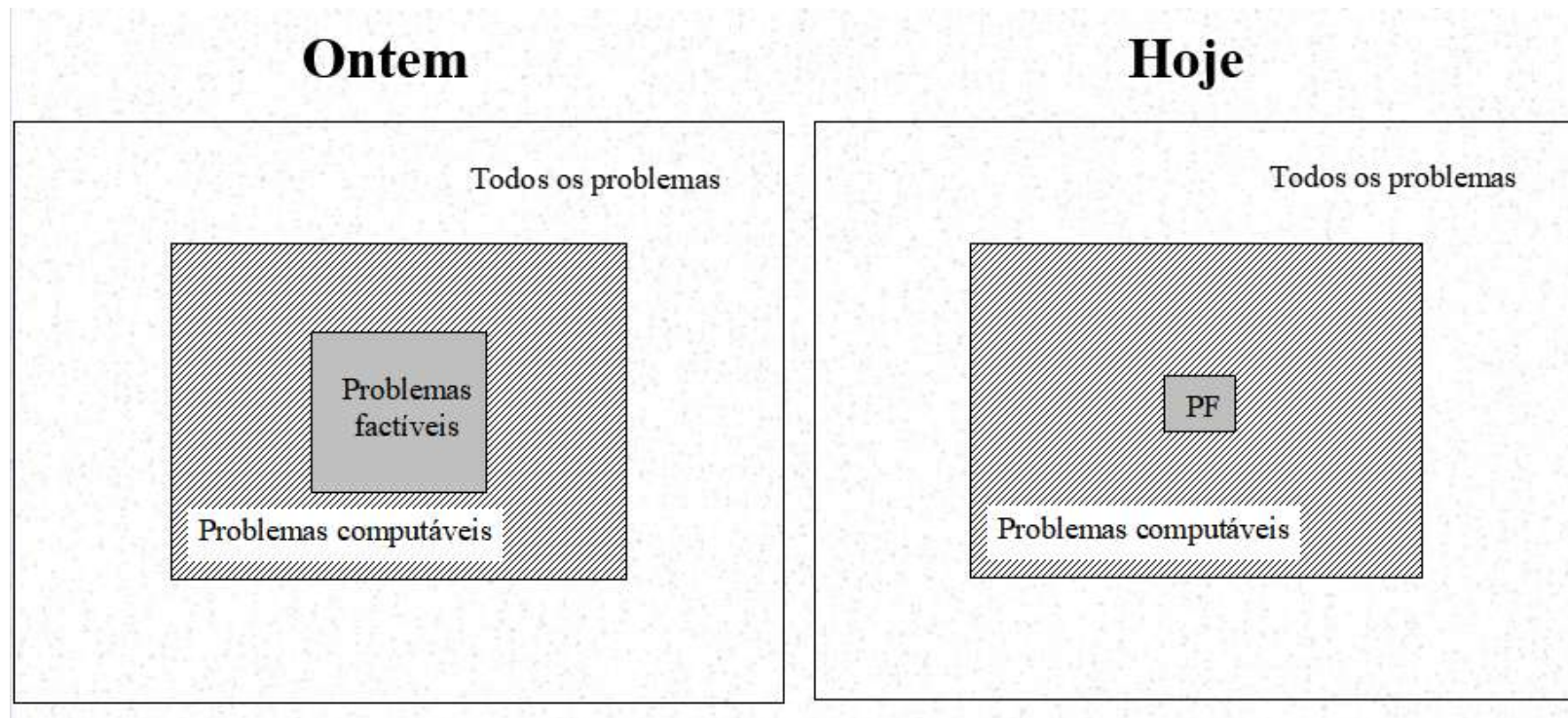


Figura 19 – Perspectiva de crescimento da quantidade de dados coletados e disponíveis, considerando três fontes de geração principais

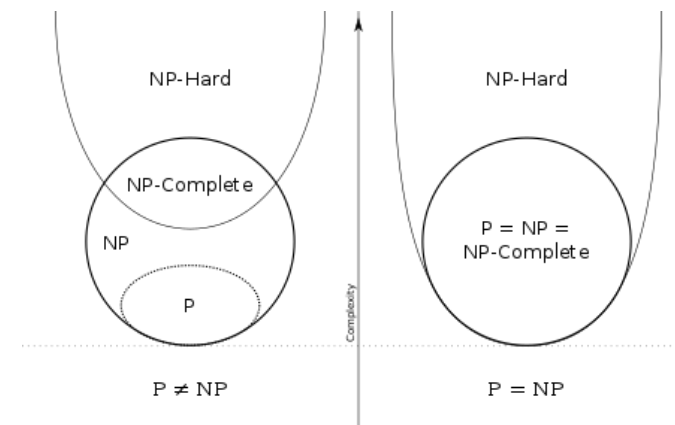
Fonte: <https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>



- Por mais que a tecnologia avance, estamos perdendo a capacidade de resolver problemas de interesse prático.
- A natureza desses problemas pode até não variar muito e o seu tamanho pode crescer apenas linearmente ou até mais lentamente que isso.

- A lei de Moore se refere ao fato de que, ao longo da história dos processadores digitais, o número de transistores nos circuitos integrados dobra aproximadamente a cada dois anos, o que representa um crescimento exponencial no poder de processamento e memória.
- A questão é que, mesmo com uma expansão linear no tamanho  $n$  de uma ampla gama de problemas computáveis de interesse prático (NP-completos e NP-difíceis), a sua solução vai requerer um aumento fatorial no poder computacional.
- Na fórmula a seguir, percebe-se que a função fatorial cresce a uma taxa muito maior que uma função exponencial de base fixa, como aquela da lei de Moore:

$$n! \cong \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$



- Reflexos da era da informação:



Lotfi A. Zadeh (1921-2017)

“À medida que a complexidade de um sistema aumenta, nossa habilidade de fazer afirmações precisas e que sejam significativas acerca deste sistema diminui até que um limiar é atingido, além do qual precisão e significância (ou relevância) tornam-se quase que características mutuamente exclusivas.” Zadeh (1973)

## **9 O poder da simulação computacional**

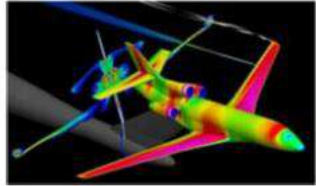

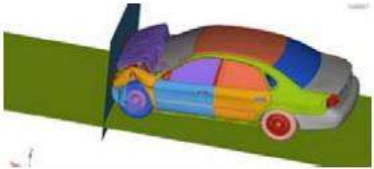

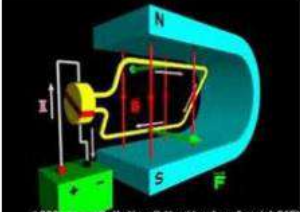
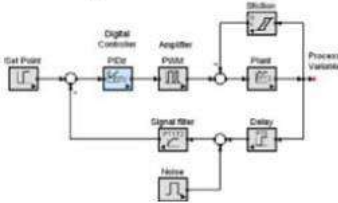
- Com a validade da lei de Moore por 5 décadas, o computador digital se tornou uma máquina de propósito geral, com uma gigantesca capacidade de simular mundos virtuais e um enorme potencial de contribuição para a ciência (e.g. na demonstração de teoremas), para a indústria (e.g. na validação de produtos) e para a sociedade (e.g. no entretenimento, na saúde e na educação).
- Há uma clara tendência de que as pessoas passem mais e mais tempo em mundos virtuais, incluindo atividades profissionais e momentos de lazer.
- A simulação computacional é de relevância central para a inteligência artificial em geral, e para o aprendizado de máquina em particular.
- Como já apresentado, as redes neurais artificiais são formadas pela interconexão de neurônios artificiais, os quais realizam mapeamentos não-lineares locais e o seu fluxo de informação pode ser representado matematicamente.
- Torna-se premente, portanto, responder a duas questões fundamentais:

- Uma simulação computacional pode ser usada como fonte de novos conhecimentos?
- Uma simulação computacional pode nos dizer alguma coisa que já não sabemos?

- Note que:

✓ Uma simulação não é melhor do que as hipóteses consideradas em sua concepção;

✓ Um computador só pode fazer o que ele é programado para fazer.

<p>Aerodynamic Simulation</p> 	<p>Mechanism and Dynamic Simulation</p> 
<p>Crash Simulation</p> 	<p>Structural Simulation</p> 
<p>Electric Motor Simulation</p> 	<p>Control System Simulation</p> 

- A resposta a essas duas questões é afirmativa. E esse “sim” é suportado pelo fato de que, mesmo havendo um controle pleno sobre as premissas que regem o mundo simulado, pode ser muito difícil antecipar suas implicações.



Herbert Simon (1916-2001)

## 10 Ser igual × Ser não-distinguível

- A inteligência artificial é uma área da computação voltada para a formalização e implementação de processamento inteligente de informação em máquinas, numa tentativa de construir mecanismos computacionais para atividades que supostamente requerem inteligência quando feitas por seres humanos.
- A denominação IA foi proposta em 1956, associada à ciência e à engenharia aplicadas na concepção de máquinas (ou organismos cibernéticos, termo cunhado em 1948) com comportamento adaptativo orientado a objetivos e capazes de automatizar tarefas associadas ao comportamento inteligente.
- É um dos pilares da Tecnologia de Informação, responsável por dotar os computadores de maior capacidade de representação do conhecimento, de inferência centrada em dados e de soluções de apoio à tomada de decisão.
- Tem como um de seus pilares a área de Inteligência Computacional, que se estabeleceu a partir de 1994, no congresso IEEE WCCI.



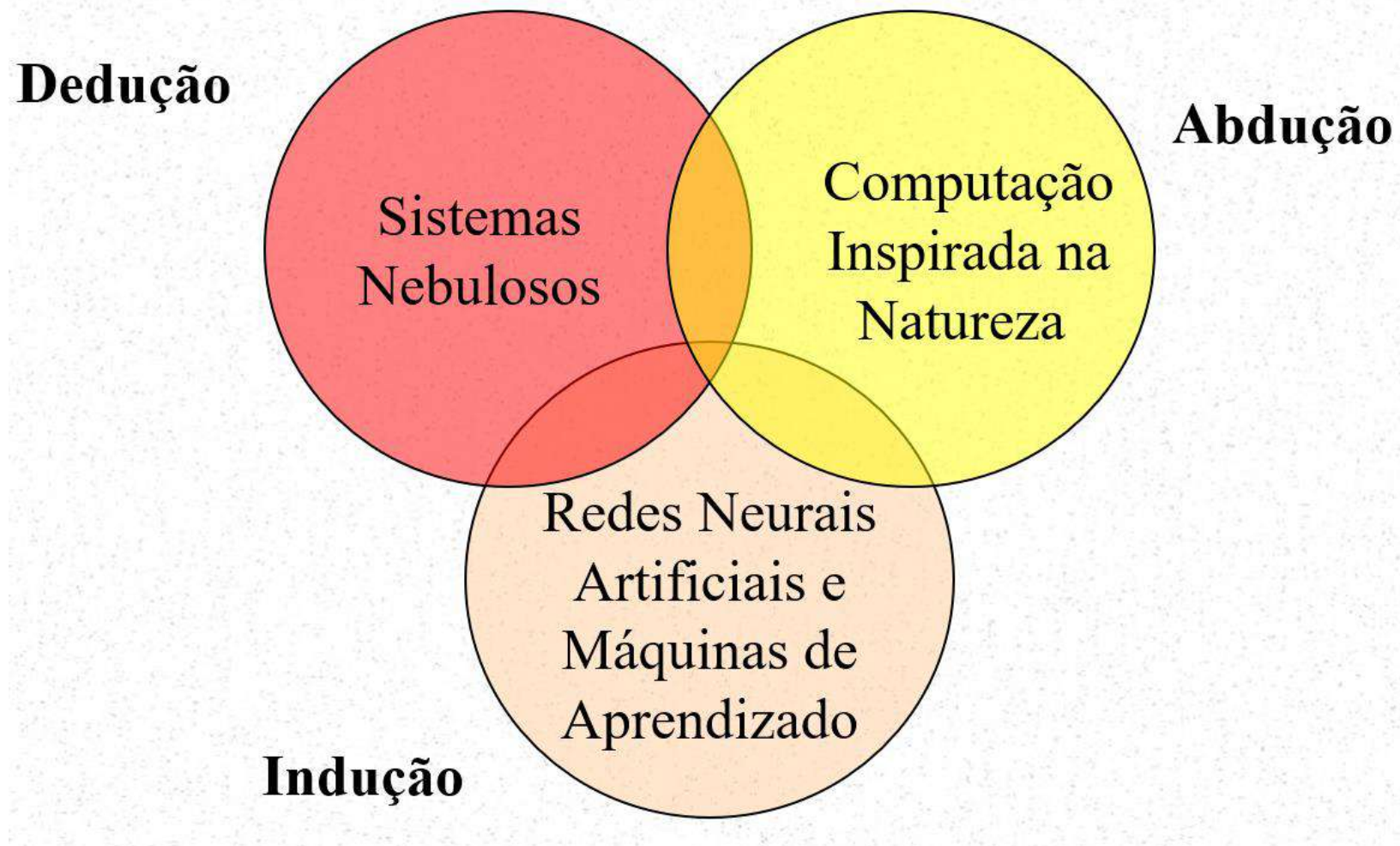


Figura 20 – As três vertentes da inteligência computacional

- De fato, o conhecimento (científico) avança ao recorrer a três tipos de inferência:
  - ✓ Dedução: A conclusão segue logicamente das premissas (está contida nas premissas). Portanto, a conclusão é lógica e sempre verdadeira. Não acrescenta conhecimento novo e corresponde ao silogismo aristotélico.
  - ✓ Indução: De juízos (enunciados) particulares se chega a um juízo (enunciado) geral. A conclusão não segue logicamente das premissas (não está contida nas premissas). Portanto, a conclusão não é lógica. Acrescenta conhecimento novo (generalização). A conclusão necessariamente consiste na extensão uniforme da evidência. Na prática, grande parte do nosso comportamento se baseia na indução, pois é através dela que ganhamos capacidade de prever o futuro.
  - ✓ Abdução: Se sustenta na seguinte sequência de ações: postular hipóteses, escolher a melhor (por exemplo, de acordo com o seu poder explicativo) e postular uma ligação entre o poder explicativo e os fatos observados.

- Vamos tomar um exemplo didático concebido para admitir os três processos de inferência. Considere as seguintes proposições:
  - ✓  $A$ : Todos os gizes da caixa são brancos.
  - ✓  $B$ : Todos os gizes da mesa vieram da caixa.
  - ✓  $C$ : Todos os gizes da mesa são brancos.
- Agora, suponha que elas são verdadeiras aos pares e infira a terceira proposição. Cada inferência vai ser exemplo de um dos três tipos mencionados:

Dedução:  $(A \wedge B) \rightarrow C$

Indução:  $(B \wedge C) \Rightarrow A$

Abdução:  $(A \wedge C) \dashv\vdash B$

- Como já indicado, o aprendizado de máquina opera com a inferência indutiva. A partir de um conjunto de dados disponíveis para treinamento, induz-se um modelo de aprendizado que deve apresentar boa capacidade de generalização.

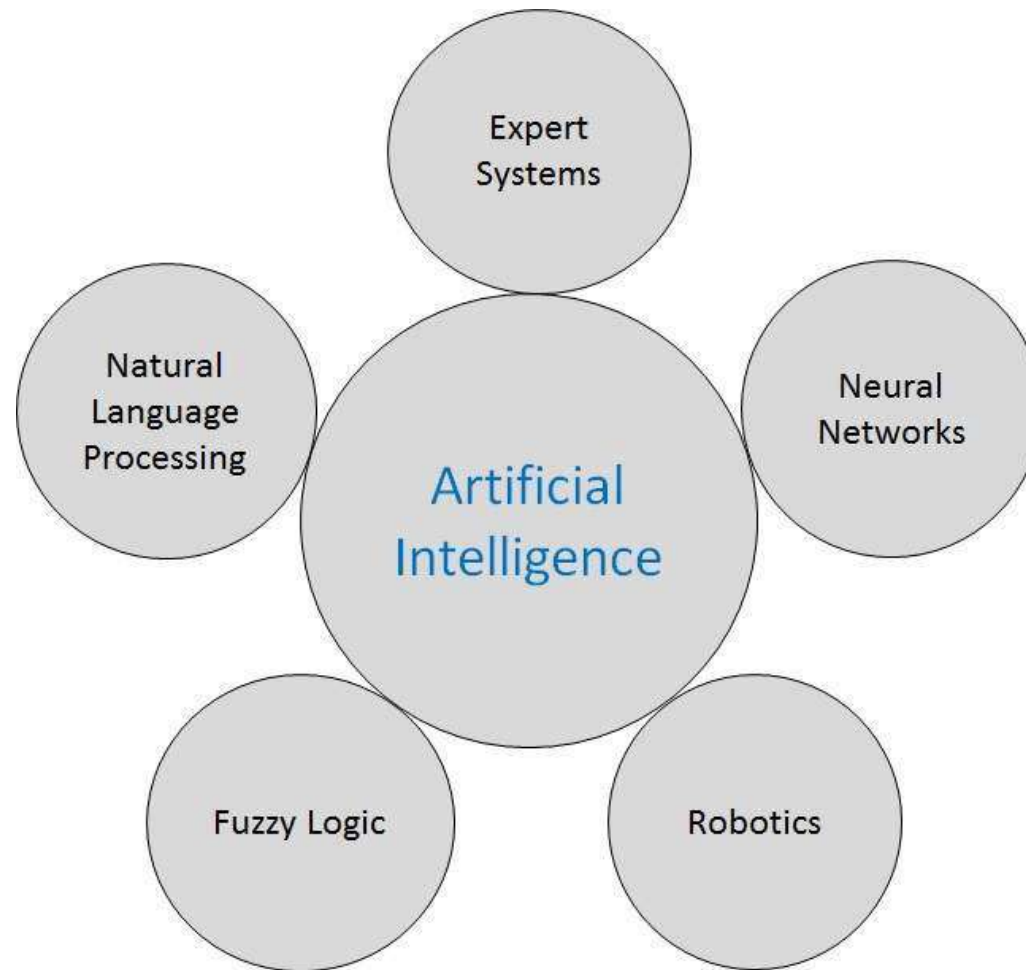


Figura 21 – Taxonomia para a área de inteligência artificial, não evidenciando a área de inferência abdutiva. IA é, de fato, mais do que esta figura mostra.

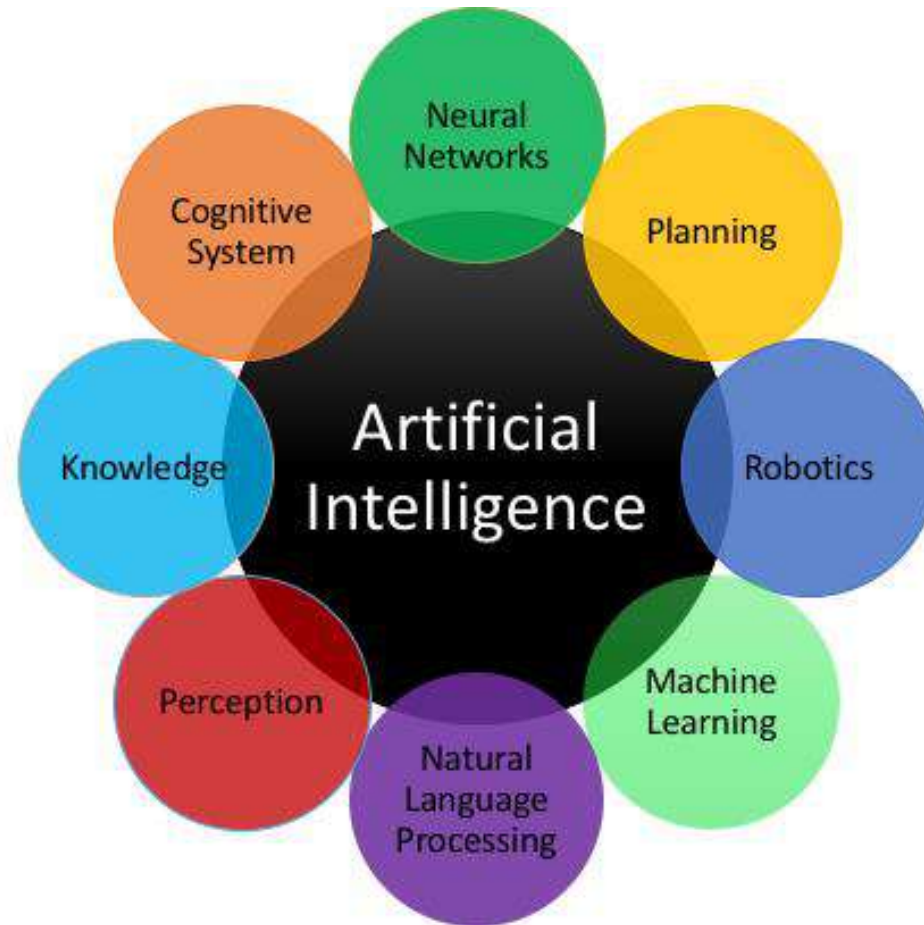


Figura 22 – Outra taxonomia para a área de inteligência artificial, ainda sem evidenciar a área de inferência abdutiva. IA é, de fato, mais do que esta figura mostra.

## 10.1 Etapas do processo de aprendizado

1. Entender os conceitos envolvidos
  2. Usar / Praticar os conceitos envolvidos
- As etapas acima são geralmente bem trabalhadas por humanos e por máquinas. Mas as etapas que vêm a seguir, menos, por humanos e por máquinas.
3. Explicar a outras pessoas os conceitos envolvidos
  4. Transpor para outros contextos
  5. Criticar / Detectar limitações
  6. Propor melhorias e extensões possíveis
  7. Substituir por uma teoria ou método distinto e com desempenho superior
- A questão que se levanta, e que representa o núcleo central desta Seção 10, é:

Até onde as máquinas que aprendem podem chegar?

De acordo com uma pesquisa de uma universidade inglesa, não importa em qual ordem as letras de uma palavra estão, a única coisa importante é que a primeira e última letras estejam no lugar certo. O resto pode ser uma qualquer que você pode ainda ler sem problema. Isso ocorre porque nós não lemos cada letra isolada, mas a palavra como um todo.

- Você consegue ler o texto acima?
- Uma máquina consegue ler o texto acima?

De acordo com uma pesquisa de uma universidade inglesa, não importa em qual ordem as letras de uma palavra estão, a única coisa importante é que a primeira e última letras estejam no lugar certo. O resto pode ser uma bagunça que você pode ainda ler sem problema. Isso ocorre porque nós não lidamos com cada letra isolada, mas a palavra como um todo.

**Tradução realizada em 13/03/2021 pelo Google Translate:**

According to a pseudose of an ignominious university, it doesn't matter in which the letters of an ethno plume are, the essential element is that the lyrical pyramid and uterus are in the correct place. The rest could be a bahu that you could read without problem. Itso occurs because we do not fight with each other, but the strike as an animal.

- Você consegue responder o que faltou para que uma máquina se desse melhor na tarefa de tradução do texto?



- Seja para o ser humano, seja para a máquina (agente de inteligência artificial), é preciso formular o problema de modo a “facilitar” a tarefa de reconhecimento de padrões, ou seja, é preciso focar na estratégia de representação do conhecimento. Veremos em módulos subsequentes do curso que *deep learning* se sobressai justamente nesta tarefa de representação do conhecimento.

## 10.2A sala chinesa de Searle

- Os computadores digitais são máquinas lógico-simbólicas, algorítmicas. A matemática envolvida é restrita, pois só trabalha com um conjunto discreto de símbolos e funções. Com isso, um programa é uma sequência de regras determinísticas sobre como transformar, transportar e armazenar dados.
- SEARLE (1991) descreve uma sala com uma pessoa que domina o idioma inglês, o operador. Muitos cestos com ideogramas chineses estão na sala, assim como um livro de regras, escrito em inglês, de como combinar os ideogramas chineses.

- O operador recebe por uma abertura de entrada (na parede da sala) uma sequência de ideogramas chineses e, consultando o livro de regras, combina esses ideogramas de entrada e alguns que estão nos cestos, compondo uma nova sequência. Esta nova sequência é então passada por uma abertura de saída (na parede da sala).
- Embora o operador não saiba, ele está respondendo a perguntas no idioma chinês.
- SEARLE (1991) argumenta que há uma diferença marcante entre este operador e uma pessoa que domina o idioma chinês e responde às mesmas perguntas sem usar o livro de regras. O primeiro está apenas seguindo regras sintáticas. O segundo está associando semântica (significado) ao que está fazendo e, portanto, está fazendo muito mais que o primeiro.
- A conclusão de SEARLE (1991) é que os computadores, por serem máquinas sintáticas, podem substituir o operador.
- Há algo mais em ter uma mente do que executar processos formais ou sintáticos.

- Logo, programas não são suficientes para atribuir mentes a computadores.
- Conclusão: Computadores nunca podem pensar, porque pensar envolve semântica. Com isso, computadores nunca poderão ser inteligentes.

### 10.30 teste de Turing

- TURING (1950) teve uma intuição de que nosso pensamento é uma atividade interior muito especial, e que seria eventualmente impossível descrever seu processo cientificamente:

“Será que as máquinas não poderiam realizar algo que deveria ser descrito como pensar mas que é muito diferente do que um ser humano faz? Esta objeção é muito forte, mas ao menos podemos dizer que se, contudo, uma máquina puder ser construída para jogar o jogo da imitação satisfatoriamente, nós não precisamos nos preocupar com essa objeção.”

- O teste de Turing afirma que um computador pode ser chamado de inteligente se ele puder enganar um ser humano ao fazê-lo acreditar que o computador é um humano. O teste consiste em um ser humano realizar um interrogatório através de um terminal, visando descobrir se do outro lado do terminal se encontra um ser humano ou um computador. Caso seja um computador que esteja fornecendo as respostas e o interrogador não conseguir decidir se se trata de um ser humano ou um computador, então este computador é dito ter sido aprovado no teste de Turing.
- Há variantes do teste de Turing, dentre as quais destaca-se:
  - ✓ O interrogador tem do outro lado do terminal um humano e um computador, sendo que o objetivo é dizer quem é o humano e quem é o computador.
- O teste de Turing é equivalente ao “jogo do fingimento”, onde existem três participantes, sendo um homem, uma mulher e um interrogador de qualquer sexo. O objetivo do jogo é o interrogador, através de perguntas aos participantes e sem

ter acesso a aspectos físicos dos entrevistados, tentar identificar quem é o homem e quem é a mulher, sendo que o homem tenta impedir que a discriminação seja realizada pelo interrogador.

- Com este teste, Turing abre mão da necessidade de definir objetivamente os atributos necessários para a caracterização de uma máquina como um sistema inteligente.
- Além disso, mesmo voltada para uma aplicação específica de IA, Turing mostrou que a inteligência artificial pode ser implementada com sucesso por meios distintos daqueles utilizados pela mente humana para realizar cognição, desde que o resultado final seja não-distinguível daquele que seria obtido por um ser realmente inteligente.
- Conclusão: Uma máquina, particularmente aquela fundamentada na computação digital, não pode ser inteligente, mas é possível fazer com que o seu comportamento não seja distinguível de um comportamento inteligente.

Um guarda noturno estava fazendo sua ronda cotidiana quando observou um homem ajoelhado junto a um poste de iluminação procurando algo. Ele então se aproximou e perguntou:

– Perdeu alguma coisa, moço?

O homem então respondeu:

– Sim, perdi um molho de chaves.

O guarda, tentando ajudar, quis saber:

– Você tem alguma ideia de onde o perdeu?

Para sua surpresa, o homem disse com naturalidade:

– Sim, foi lá no fim da rua, no escuro.

Sem entender mais nada, o guarda insistiu:

– Mas então o que você faz aqui, procurando embaixo deste poste de iluminação?

E então o homem explicou:

– É que aqui é o único lugar onde existe luz suficiente para eu poder encontrá-lo.

## 11 Uma excursão por *deep learning*

### 1. Livro-texto:

- ✓ CHOLLET, F. “Deep Learning with Python”, Manning Publications, 2017.
- ✓ GOODFELLOW, I.; BENGIO, Y. & COURVILLE, A. “Deep Learning”, The MIT Press, 2016. ( <http://www.deeplearningbook.org/> )
- ✓ GULLI, A. & PAL, S. “Deep Learning with Keras: Implementing deep learning models and neural networks with the power of Python”, Packt Pub., 2017.

### 2. Livro online:

- ✓ NIELSEN, M. “Neural Networks and Deep Learning”, 2012-2019. ( <http://neuralnetworksanddeeplearning.com/> )
- ✓ Deep Learning Book Brasil, 2018. ( <http://deeplearningbook.com.br/> )

### 3. Livro de apoio:

- ✓ RASCHKA, S. “Python Machine Learning”, Packt Publishing Ltd., 2015.  
(destaque para Chapter 13)

#### 4. Papers marcantes + reviews:

- ✓ <https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>
- ✓ LECUN, Y.; BENGIO, Y. & HINTON, G. “Deep learning”, Nature, vol. 521, pp. 436-444, 28 May 2015.
- ✓ BENGIO, Y.; COURVILLE, A. & VINCENT, P. “Representation Learning: A Review and New Perspectives”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, 2013.
- ✓ BENGIO, Y. “Learning Deep Architectures for AI”, Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.
- ✓ SCHMIDHUBER, J. “Deep learning in neural networks: An overview”, Neural Networks, vol. 61, pp. 85-117, 2015.



- ✓ DENG, L. “Three Classes of Deep Learning Architectures and Their Applications: A Tutorial Survey”, APSIPA Transactions on Signal and Information Processing, 2012.
- ✓ VINYALS, O., TOSHEV, A., BENGIO, S. & ERHAN, D. “Show and tell: A neural image caption generator”, arXiv:1411.4555v2, 2015.
- ✓ XU, K., BA, J., KIROS, R., COURVILLE, A., SALAKHUTDINOV, R., ZEMEL, R. & BENGIO, Y. “Show, attend and tell: Neural image caption generation with visual attention”, arXiv:1502.03044v3, 2016.

## 5. Vídeos / apresentações:

- ✓ [https://www.youtube.com/watch?v=\\_1Cyyt-4-n8](https://www.youtube.com/watch?v=_1Cyyt-4-n8)
- ✓ Tutorial proferido por Geoffrey Hinton, Yoshua Bengio & Yann LeCun [ <http://www.iro.umontreal.ca/~bengioy/talks/DL-Tutorial-NIPS2015.pdf> ].
- ✓ E muito mais cursos e tutoriais ...

## 6. Cursos acadêmicos:

- ✓ Aprendizado de máquina – Andrew Ng – Stanford University (Coursera)  
[ <https://www.coursera.org/learn/machine-learning> ]
- ✓ Redes neurais para aprendizado de máquina – Geoffrey Hinton – University of Toronto [ <https://www.coursera.org/learn/neural-networks> ]
- ✓ Aprendizado de máquina – Nando de Freitas – Oxford University  
[ <https://www.cs.ox.ac.uk/people/nando.defreitas/machinelearning/> ]
- ✓ Redes neurais artificiais – Hugo Larochelle – Université de Sherbrooke  
[ [http://info.usherbrooke.ca/hlarochelle/neural\\_networks/content.html](http://info.usherbrooke.ca/hlarochelle/neural_networks/content.html) ]
- ✓ Redes convolucionais para reconhecimento de padrões em imagens – Andrej Karpathy (versão 2016) – Stanford University [ <http://cs231n.stanford.edu/> ]
- ✓ Processamento de linguagem natural com *deep learning* – Christopher Manning – Stanford University [ <http://web.stanford.edu/class/cs224n/> ] e David Socher – Stanford University [ <http://cs224d.stanford.edu/> ]

## 7. Escolha de hiperparâmetros:

- ✓ How do we know how many layers to use, how many conv layers, what are the filter sizes, or the values for stride and padding? These are not trivial questions and there isn't a set standard that is used by all researchers. This is because the network will largely depend on the type of data that you have. Data can vary by size, complexity of the image, type of image processing task, and more. When looking at your dataset, one way to think about how to choose the hyperparameters is to find the right combination that creates abstractions of the image at a proper scale. ( <https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/> )
- ✓ BERGSTRA, J., BARDENET, R., BENGIO, Y. & KÉGL, B. “Algorithms for Hyper-Parameter Optimization”, Advances in Neural Information Processing Systems 24 (NIPS'2011), 9 pages, 2011.

## 8. Transfer learning:

- ✓ Mesmo não dispondo de “zilhões” de amostras, é possível produzir redes neurais com arquiteturas profundas de alto desempenho, adotando na etapa de pretraining os extratores de atributos de redes neurais treinadas (por outros desenvolvedores) com “zilhões” de amostras e foco de aplicação distinto.

## 9. Outros conceitos relevantes:

- ✓ ReLU (e suas extensões)
- ✓ Mini-batch: LI, M., ZHANG, T., CHEN, Y. & SMOLA, A.J. “Efficient Mini-batch Training for Stochastic Optimization”, KDD’2014.
- ✓ Manifold (high-dimensional data lies in a low-dimensional manifold)
- ✓ Softmax ( [https://en.wikipedia.org/wiki/Softmax\\_function](https://en.wikipedia.org/wiki/Softmax_function) )
- ✓ Cross-entropy: JANOCHA, K. & CZARNECKI, W.M. “On Loss Functions for Deep Neural Networks in Classification”, *Schedae Informaticae*, vol. 25, pp. 49-59, 2016.

## **10. Data augmentation:**

- ✓ Estratégia para evitar overfitting: Wu, R., Yan, S., Shan, Y., Dang, Q. & Sun, G. “Deep Image: Scaling up Image Recognition”, arXiv:1501.02876v2, 2015.

## **11. Dropout:**

- ✓ Estratégia para evitar overfitting: SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. & SALAKHUTDINOV, R. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014.

## **12. Autoencoder (pretraining):**

- ✓ VINCENT, P.; LAROCHELLE, H.; LAJOIE, I.; BENGIO, Y. & MANZAGOL, P.-A. “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion”, Journal of Machine Learning Research, vol. 11, pp. 3371-3408, 2010.

### **13. Outras técnicas que contribuem para a regularização e o aprendizado:**

- ✓ Max-norm regularization (permite taxas de aprendizado mais altas)
- ✓ Normalização / Escalamento da saída de cada camada intermediária
- ✓ Weight sharing
- ✓ Pooling | Downsampling

### **14. Métodos de otimização para o ajuste de pesos:**

- ✓ LE, Q.V., NGIAM, J., COATES, A., LAHIRI, A., PROCHNOW, B., NG, A.Y. “On optimization methods for deep learning”, Proceedings of the 28th International Conference on Machine Learning (ICML’2011), pp. 265-272, 2011.
- ✓ PASCANU, R. & BENGIO, Y. “Revisiting natural gradient for deep networks”, arXiv:1301.3584v7, 2014.

## 15. Camadas convolucionais:

- ✓ ZEILER, M.D. & FERGUS, R. “Visualizing and Understanding Convolutional Networks”, Proceedings of the European Conference on Computer Vision (ECCV’2014), Lecture Notes on Computer Science, Springer, vol. 8689, pp. 818-833, 2014.
- ✓ <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/>

## 16. Deep Recurrent Neural Networks

- ✓ GREFF, K., SRIVASTAVA, R.K., KOUTNÍK, J., STEUNEBRINK, B.R. & SCHMIDHUBER, J. “LSTM: A Search Space Odyssey”, IEEE Transactions on Neural Networks and Learning Systems, to appear, 2017.

- ✓ VISIN, F., KASTNER, K., CHO, K., MATTEUCCI, M., COURVILLE, A. & BENGIO, Y. “ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks”, arXiv:1505.00393v3, 2015.
- ✓ CHO, K., VAN MERRIËNBOER, B., GULCEHRE, C., BAHDANAU, D., BOUGARES, F., SCHWENK, H., BENGIO, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, arXiv:1406.1078v3, 2014.
- ✓ Christopher Olah [ <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> ]
- ✓ Andrej Karpathy [ <http://karpathy.github.io/2015/05/21/rnn-effectiveness/> ]

## 17. Attention models:

- ✓ BA, J., MNIH, V. & KAVUKCUOGLU, K. “Multiple object recognition with visual attention”, arXiv:1412.7755v2, 2015.



## 18. Generative Adversarial Nets:

- ✓ GOODFELLOW, I.J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIRY, S., COURVILLE, A., BENGIO, Y. “Generative Adversarial Nets”, arXiv:1406.2661v1, 2014.
- ✓ RADFORD A.; METZ, L.; CHINTALA, S “Unsupervised representation learning with deep convolutional generative adversarial networks”, arXiv:1511.06434v2, 2016.
- ✓ CHEN, X., DUAN, Y., HOUTHOOFT, R., SCHULMAN, J., SUTSKEVER, I. & ABBEEL, P. “InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets”, arXiv:1606.03657v1, 2016.
- ✓ <http://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/>

## 19. Restricted Boltzmann machines + Deep Belief Networks:

- ✓ FISCHER, A. & IGEL, C. “An Introduction to Restricted Boltzmann Machines”, Proceedings of the 17th Iberoamerican Congress on Pattern Recognition (CIARP’2012), Lecture Notes on Computer Science, Springer, vol. 7441, pp. 14-36, 2012.
- ✓ <http://www.cs.toronto.edu/~hinton/deeprefs.html>
- ✓ <https://deeplearning4j.org/restrictedboltzmannmachine>

## 20. Deep Reinforcement Learning:

- ✓ LI, Y. “Deep Reinforcement Learning: An Overview”, arXiv:1701.07274v2, 2017.
- ✓ [http://hunch.net/~beygel/deep\\_rl\\_tutorial.pdf](http://hunch.net/~beygel/deep_rl_tutorial.pdf)
- ✓ Gorila (General Reinforcement Learning Architecture)
- ✓ Games e sistemas de recomendação

## 21. Ambientes de programação para *deep learning*:

- ✓ TensorFlow e Pytorch representam as duas bibliotecas de código aberto mais populares para *deep learning*.
- ✓ TensorFlow foi desenvolvido pelo Google e é adotado em várias soluções internas, como no sistema de reconhecimento de fala, no Gmail e no sistema de busca.
- ✓ PyTorch está sendo desenvolvido por pesquisadores da NVIDIA e de importantes universidades, como Stanford, Oxford e ParisTech. Twitter e Facebook usam PyTorch.
- ✓ Theano é outra biblioteca de código aberto para *deep learning*, muito similar ao TensorFlow.
- ✓ Keras é outra biblioteca para modelos de *deep learning*, podendo encapsular códigos de Theano e TensorFlow e, assim, implementar modelos complexos em poucas linhas de código.

- ✓ [https://en.wikipedia.org/wiki/Comparison\\_of\\_deep\\_learning\\_software](https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software)
- ✓ [https://en.wikipedia.org/wiki/Deep\\_learning#Software\\_libraries](https://en.wikipedia.org/wiki/Deep_learning#Software_libraries)

## 22. Casos de estudo que evidenciam o poder de *deep learning*:

- ✓ No link [<https://www.quora.com/How-do-I-learn-deep-learning-in-2-months>], foram apresentadas frentes de aplicação que podem ser tomadas como casos de estudo (há hiperlinks de fácil acesso):

- As is tradition, start with classifying the [MNIST dataset](#)
- Try face detection and classification on [ImageNet](#). If you are up to it, do the [ImageNet Challenge 2016](#).
- Do a Twitter sentiment analysis using [RNNs](#) or [CNNs](#)
- Teach neural networks to reproduce the artistic style of famous painters ([A Neural Algorithm of Artistic Style](#))
- [Compose Music With Recurrent Neural Networks](#)
- [Play ping-pong using Deep Reinforcement Learning](#)
- Use [Neural Networks to Rate a selfie](#)
- Automatically [color Black & White pictures using Deep Learning](#)

## 12 Referências bibliográficas

- RUSSELL, S. & NORVIG, P. “Artificial Intelligence – A Modern Approach”, 3rd. edition, Prentice Hall, 2009.
- SEARLE, J.R. “Minds, Brains and Science – the 1984 Reith Lectures”, Penguin Books, 1991.
- TURING, A.M. “Computing machinery and intelligence”, In *Mind – A Quarterly Review of Psychology and Philosophy*, Vol. LIX No. 236, pp. 433-460, Oct. 1950.