

Deep Learning Book

Capítulo 1 - Introdução

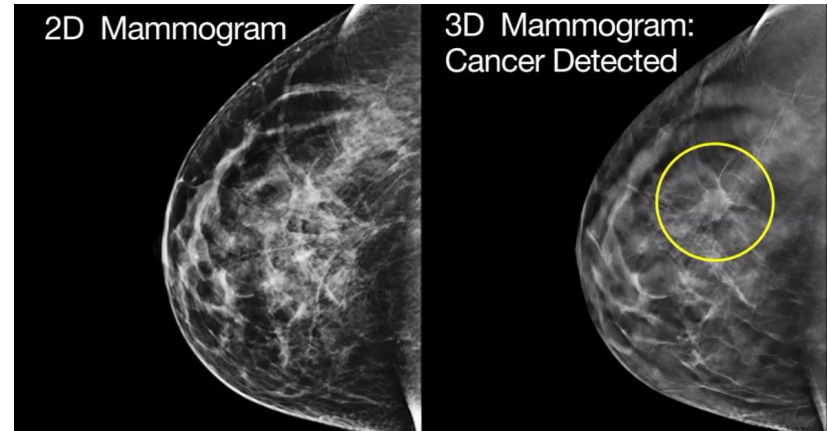
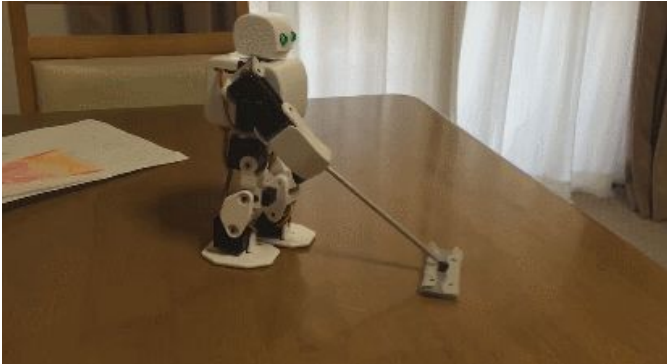
Darlan Murilo Nakamura de Araújo
darlannakamura@hotmail.com

Por que queremos IA?

Introdução

Por que queremos IA?

- Queremos softwares inteligentes para automatizar trabalhos rotineiros, entender fala ou imagens, fazer diagnósticos na medicina e auxiliar a pesquisa básica científica.



Introdução

- Os computadores passaram a fazer tarefas que eram intelectualmente difíceis para nós;



Introdução

- Mas ela falhava quando tentava realizar tarefas simples para nós, como reconhecer uma pessoa ou entender palavras faladas, ou seja, tarefas que fazemos diariamente.
- Durante muitos anos, o **verdadeiro desafio da IA** era resolver esses problemas, que eram simples para nós mas difíceis de descrever **formalmente**.

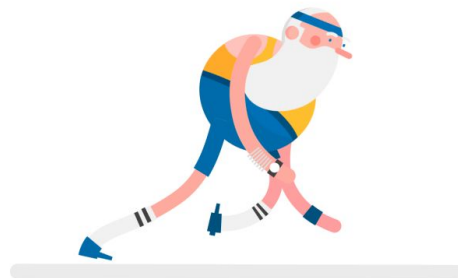
➤ Goodfellow descreve o livro como:

“ Este livro é sobre uma solução que permite os computadores aprenderem através da experiência e entendam o mundo através de uma hierarquia de conceitos, **dispensando a necessidade de um operador humano especificar formalmente todo o conhecimento.**” tradução própria

Foco do Livro

➤ Goodfellow descreve o livro como:

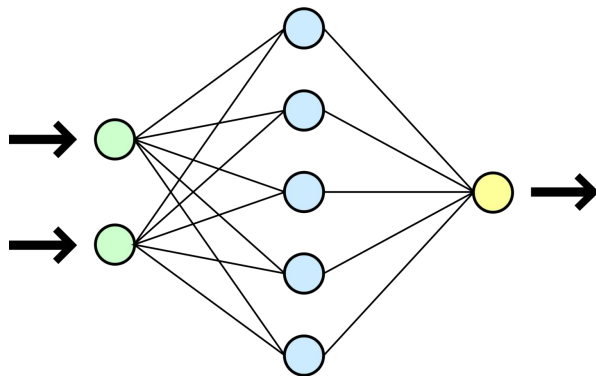
“ Este livro é sobre uma solução que permite os computadores **aprenderem através da experiência** e entendam o mundo através de uma hierarquia de conceitos, dispensando a **necessidade de um operador humano especificar formalmente todo o conhecimento.**” tradução própria



Foco do Livro

➤ Goodfellow descreve o livro como:

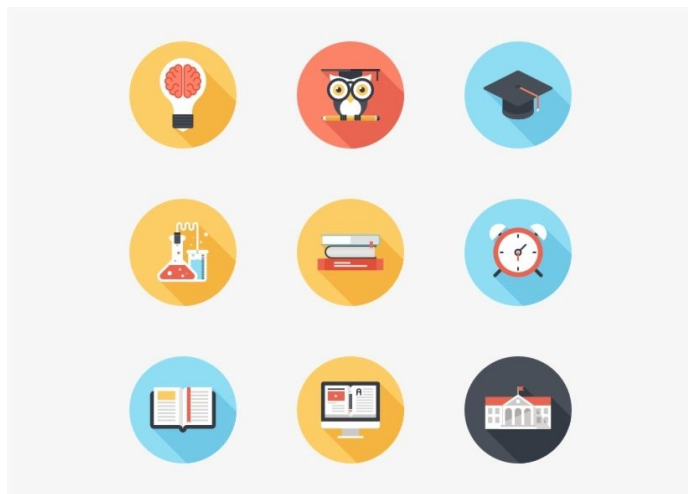
“ Este livro é sobre uma solução que permite os computadores aprenderem através da experiência e entendam o mundo através de uma **hierarquia de conceitos**, dispensando a necessidade de um operador humano especificar formalmente todo o conhecimento.” tradução própria



Foco do Livro

➤ Goodfellow descreve o livro como:

“ Este livro é sobre uma solução que permite os computadores aprenderem através da experiência e entendam o mundo através de uma hierarquia de conceitos, dispensando a necessidade de um operador humano especificar formalmente todo o conhecimento.” tradução própria



Deep Blue

IBM Deep Blue VS Garry Kasparov

Primeira partida em 1996 na Philadelphia:

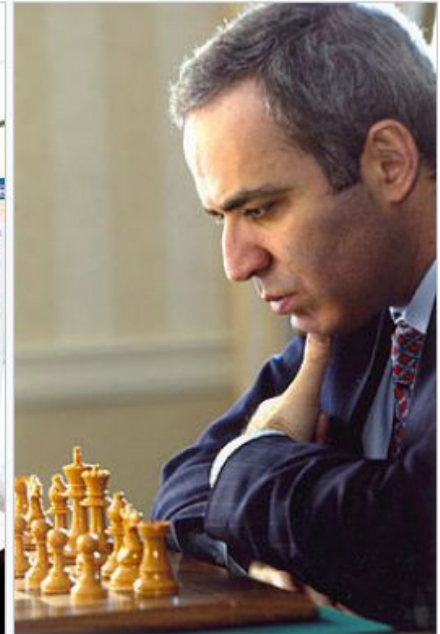
- 4 a 2 para Kasparov



Deep Blue vs. Kasparov



Deep Blue
IBM chess computer



Garry Kasparov
World Chess Champion

Deep Blue

IBM Deep Blue VS Garry Kasparov

Primeira partida em 1996 na Philadelphia:

- 4 a 2 para Kasparov

Segunda partida em 1997 em Nova York:

- 3.5 a 2.5 para Deep Blue



Deep Blue vs. Kasparov



Deep Blue
IBM chess computer



Garry Kasparov
World Chess Champion

O mundo do Xadrez

O mundo do **xadrez** é um mundo muito mais simples do que o mundo real, que pode ser descrito por uma lista de regras formais.



O mundo do Xadrez

O mundo do **xadrez** é um mundo muito mais simples do que o mundo real, que pode ser descrito por uma lista de regras formais.

Ironicamente, abstrair e formalizar regras são uma das coisas mais difíceis para seres humanos.



O mundo do Xadrez

O mundo do **xadrez** é um mundo muito mais simples do que o mundo real, que pode ser descrito por uma lista de regras formais.

Ironicamente, abstrair e formalizar regras são uma das coisas mais difíceis para seres humanos. Apenas recentemente os computadores conseguiram reconhecer objetos ou falas, coisa que qualquer ser humano mediano faz, mas em 1997 derrotou o melhor do mundo no xadrez.



Abordagem por Base de Conhecimento

Um computador pode raciocinar sobre sentenças através de linguagens formais usando **regras de inferência lógica**.

Um dos projetos mais famosos foi o Cyc (1989)

Abordagem por Base de Conhecimento

Um computador pode raciocinar sobre sentenças através de linguagens formais usando **regras de inferência lógica**.

Um dos projetos mais famosos foi o Cyc (1989)

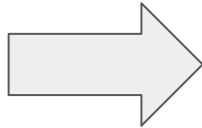
Cyc detectou uma inconsistência na história de Fred se barbeando de manhã



As dificuldades encontradas por sistemas baseados em conhecimentos *hand-coded* sugere que sistemas de IA precisam adquirir seu próprio conhecimento **extraíndo padrões dos dados**.

Esta capacidade é conhecida como aprendizado de máquina.

➤ NAIVE BAYES



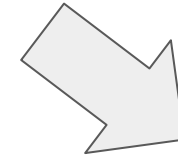
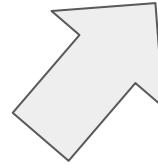
$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Labels for the equation:

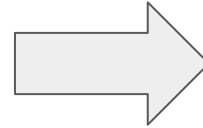
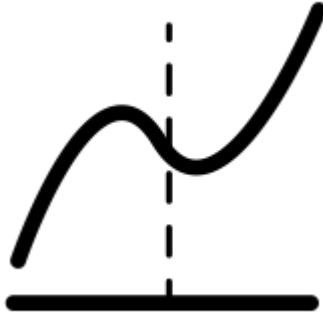
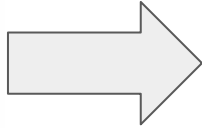
- Likelihood: $P(x|c)$
- Class Prior Probability: $P(c)$
- Posterior Probability: $P(c|x)$
- Predictor Prior Probability: $P(x)$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

NAIVE BAYES



➤ REGRESSÃO LOGÍSTICA



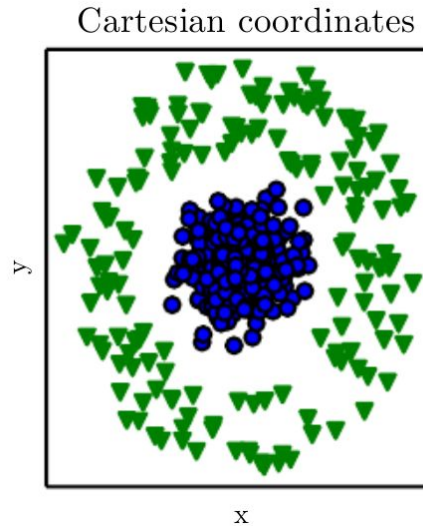
DADOS DA GRÁVIDA

REGRESSÃO LOGÍSTICA

CESÁRIA

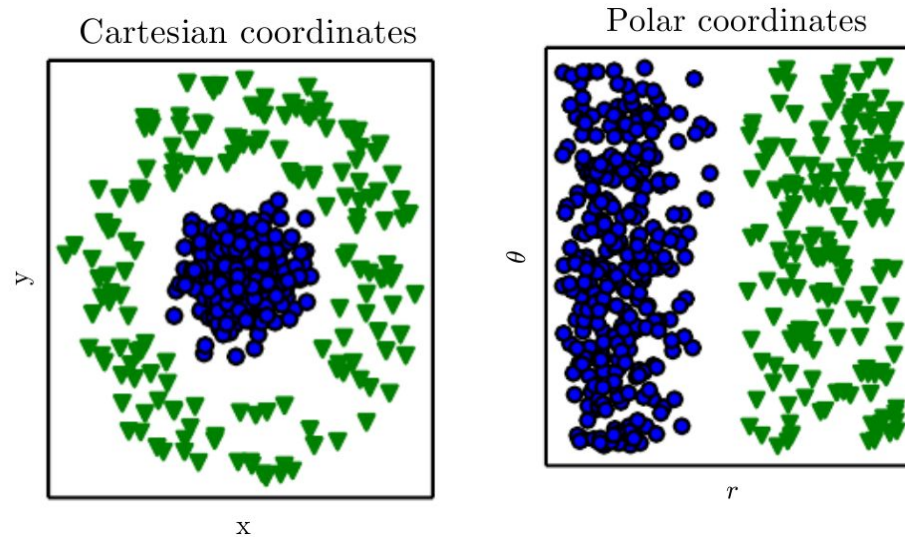
Aprendizado de Máquina

- Tudo depende de como os dados estão estruturados.
- Como você separaria os dois conjuntos?



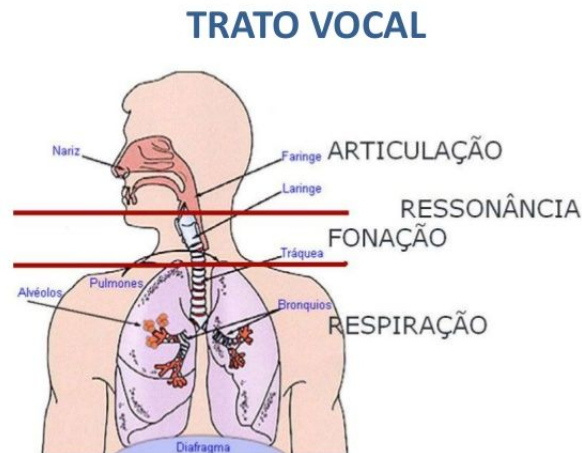
Aprendizado de Máquina

- Tudo depende de como os dados estão estruturados.
- Como você separaria os dois conjuntos?
- Às vezes é necessário enxergar as coisas de maneiras diferentes.



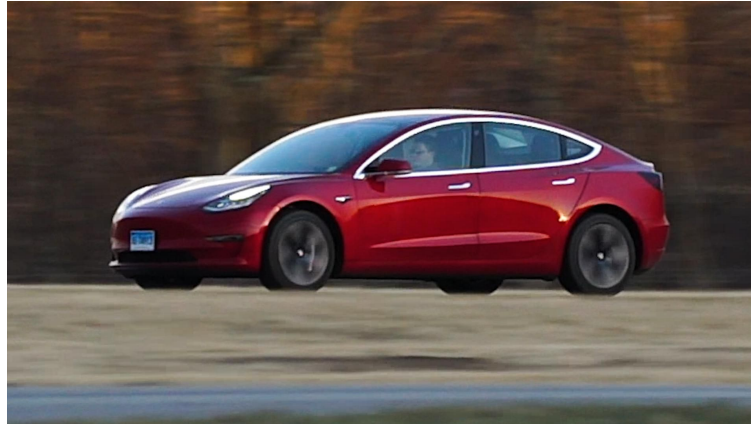
Aprendizado de Máquina

- Os algoritmos de aprendizado de máquina dependem do conjunto correto de características,
- Exemplo: uma característica muito importante para identificar quem está falando é pelo tamanho do **trato vocal**. Assim temos uma pista interessante se é um homem, uma mulher ou uma criança



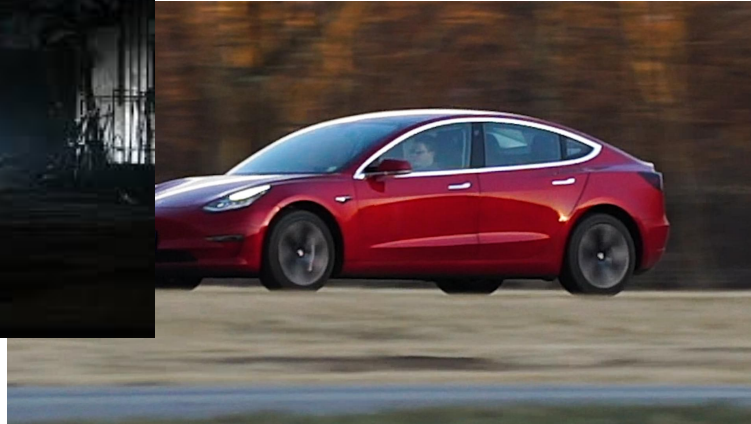
Aprendizado de Máquina

- O problema é que às vezes é difícil saber quais características devem ser extraídas. Suponha que queremos fazer um programa para **detectar carros** em fotografias.



Aprendizado de Máquina

- O problema é que às vezes é difícil saber quais características devem ser extraídas. Suponha que queremos fazer um programa para **detectar carros** em fotografias.



Aprendizado de Máquina

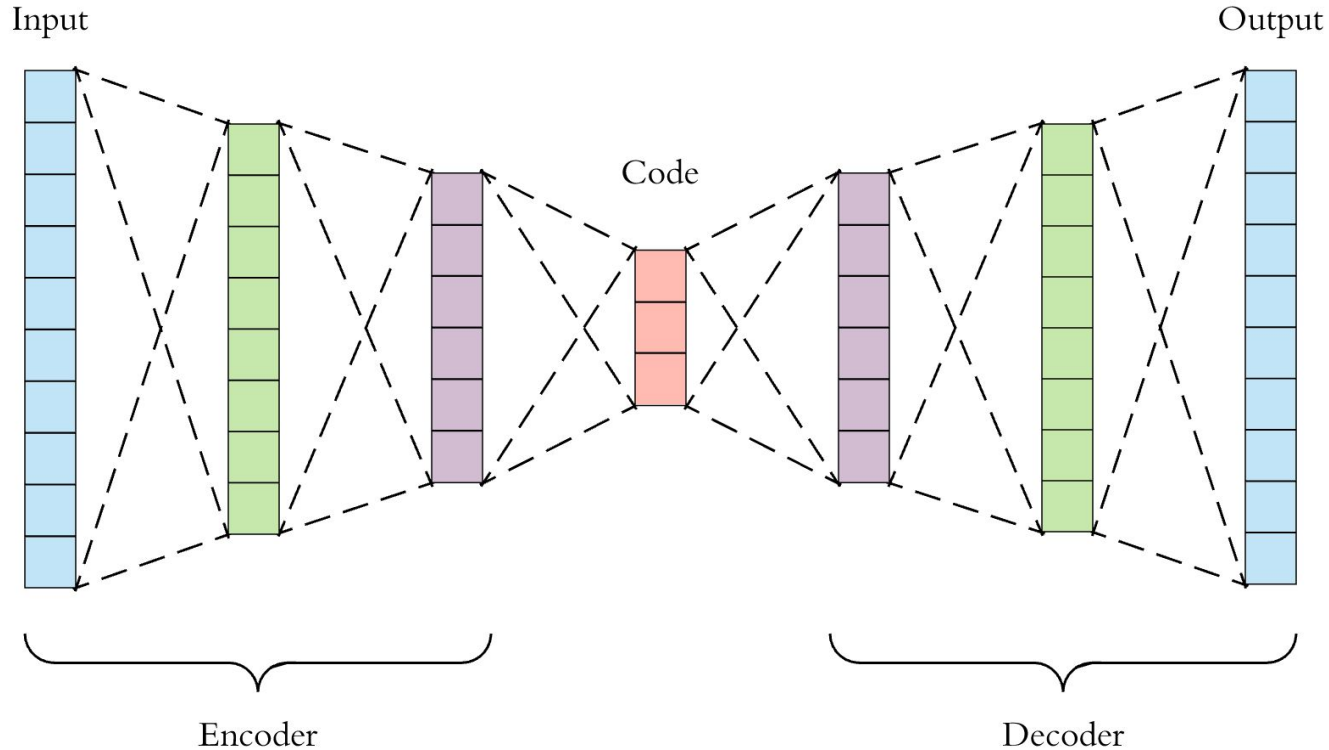
Uma solução é usar aprendizado de máquina não só para mapear da representação para a saída, mas também representar a si mesmo. Esta abordagem é chamada de **representation learning** (ou aprendizado de representação).

Aprendizado de Máquina

Uma solução é usar aprendizado de máquina não só para mapear da representação para a saída, mas também representar a si mesmo. Esta abordagem é chamada de **representation learning** (ou aprendizado de representação).

- Melhor que feito por um ser humano;
- Sistemas se adaptam a outras tarefas;
- escolher as features manualmente requer muito tempo - pode levar décadas;

➤ AUTOENCODER



Aprendizado de Representação

- Quando desenvolvemos características ou um algoritmo para aprendê-las, nosso objetivo é separar os **fatores de variação** que explicam os dados observados.
- Os **fatores de variação** são conceitos ou abstrações dos dados.
- Exemplo:
- **ÁUDIO**: idade de quem fala, sexo, sotaque, etc.
- **IMAGEM DE UM CARRO**: posição do carro, cor, ângulo, brilho, etc.

Aprendizado de Representação

- O problema é que **fatores de variação** influenciam muito nos dados observados
- um pixel de um carro vermelho pode ser muito parecido com o de um preto à noite, a silhueta de um carro depende do ângulo em que a foto foi tirada, etc.
- É muito difícil extrair essas features abstratas dos dados brutos.

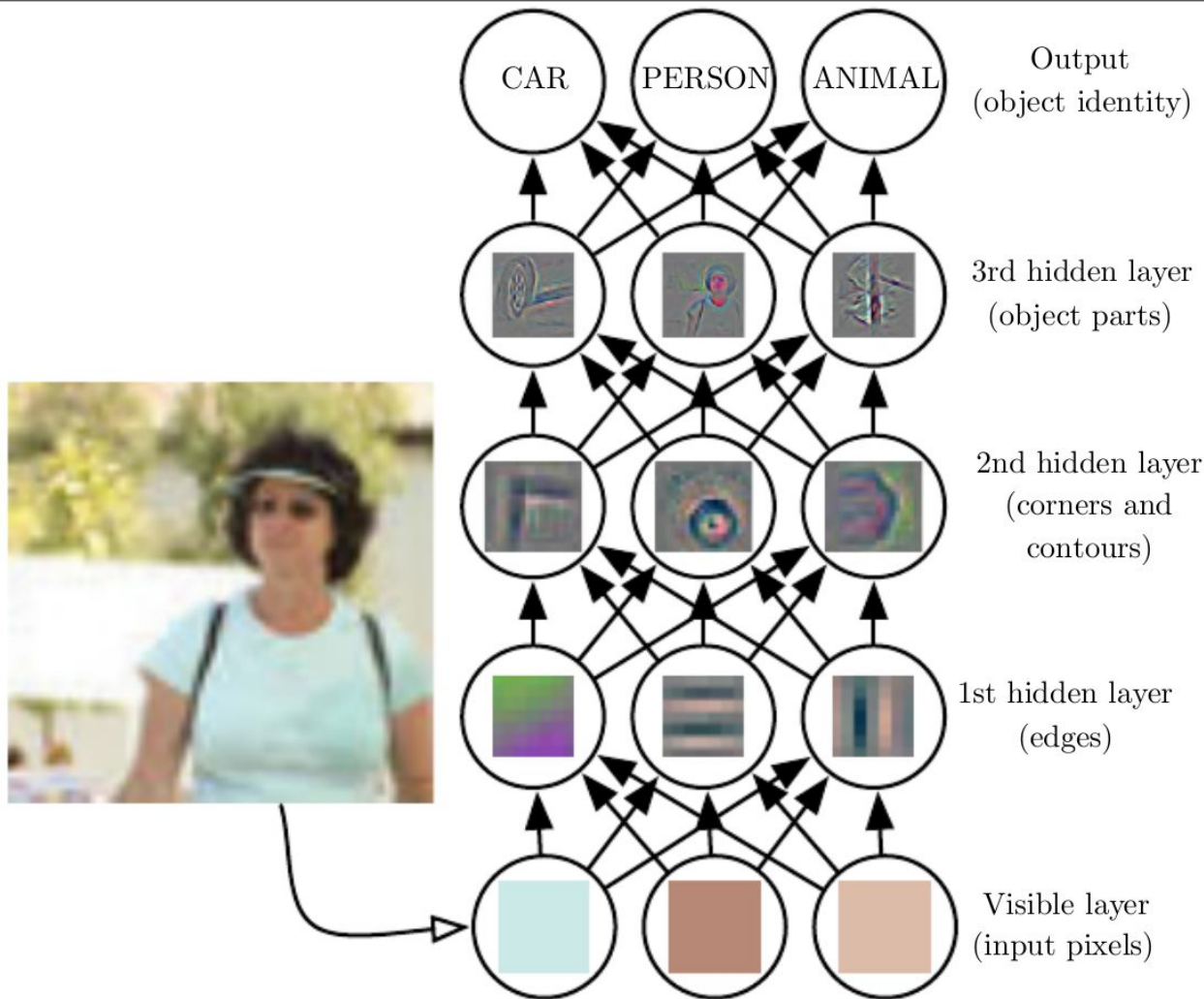
Quando é tão difícil extrair as representações quanto resolver o problema, aprendizado de representação não é mais interessante.

Deep learning resolve este problema central pois as representações são expressas em termos de outras representações mais simples.

- É possível construir conceitos complexos através de conceitos simples.
- Deep Learning é uma abordagem de IA, assim como a simbólica (fuzzy), agentes, bioinspirada, etc.

Deep Learning

Deep learning
em termos
➤ É possível
➤ Deep Learning
agentes



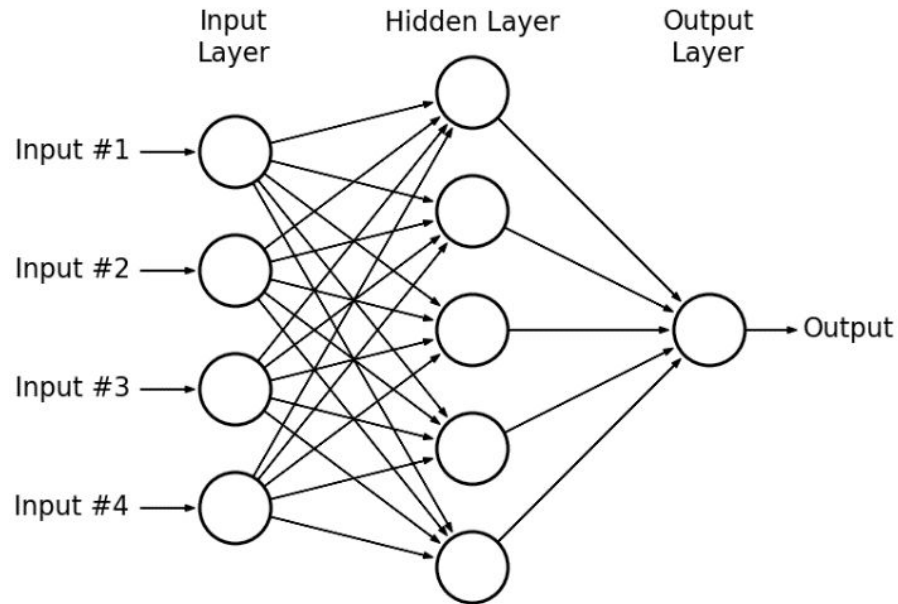
expressas

ES.

),

Multi Layer Perceptron

Função matemática que mapeia um conjunto de valores de entrada para valores de saída.



Há duas perspectivas em relação a Deep Learning:

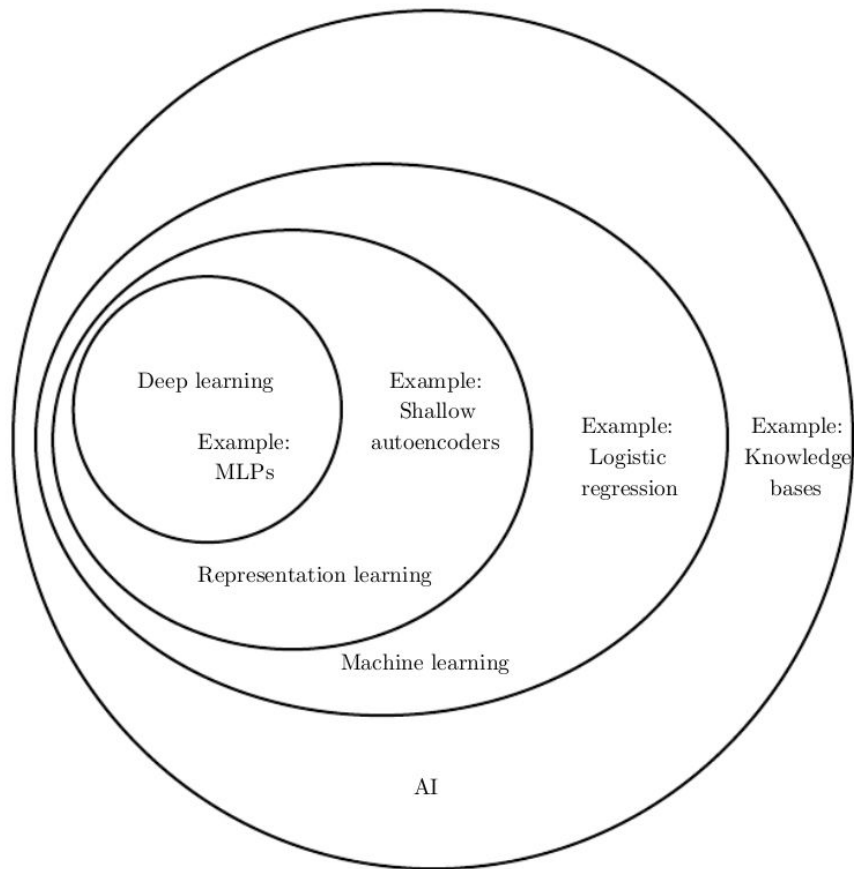
- Enxergar que o modelo irá aprender a correta representação dos dados - ou seja, toda camada é expressa e identifica um padrão a partir da entrada, mesmo que não possamos identificar o quê;

Há duas perspectivas em relação a Deep Learning:

- Enxergar que o modelo irá aprender a correta representação dos dados - ou seja, toda camada é expressa e identifica um padrão a partir da entrada, mesmo que não possamos identificar o quê;
- Podemos enxergar uma camada igual a um estado de uma célula de memória. O estado da informação ser análogo a um contador - ajuda o modelo a organizar o processo e não necessariamente representa algo a partir da entrada.

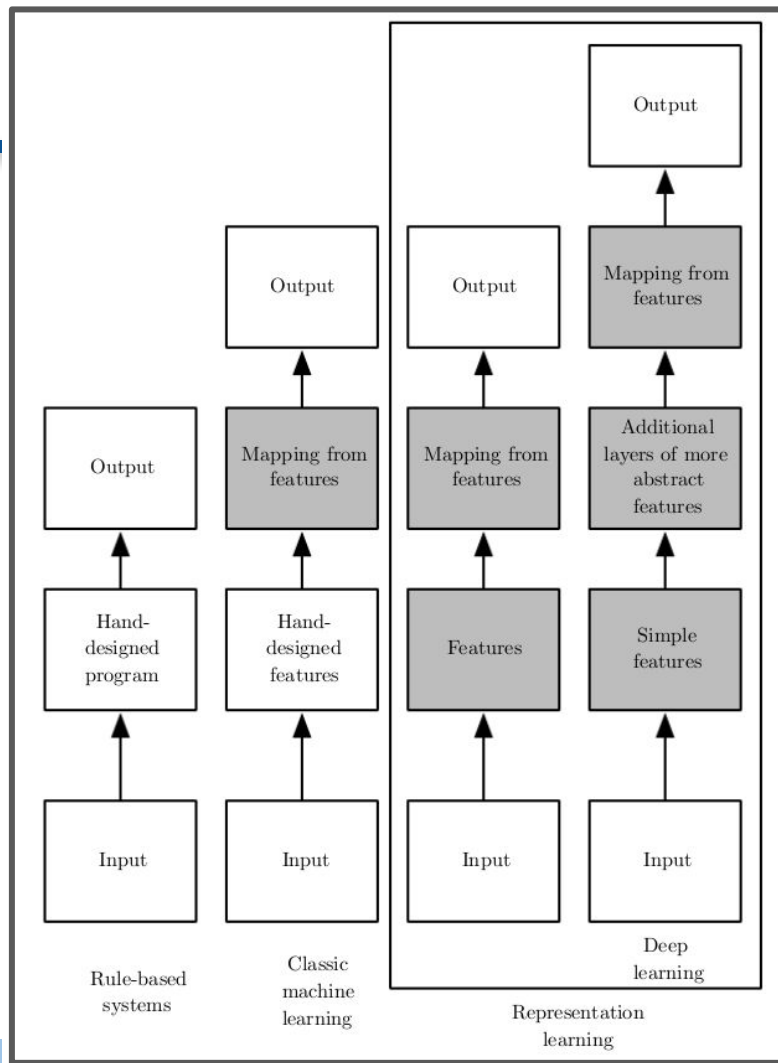
Como se relaciona Inteligência Artificial, Aprendizado de Máquina, Aprendizado de Representação e Aprendizado Profundo?

Relação entre IA, ML, RL e DL



Comparativo

Comparativo entre sistemas baseados em regras, Machine Learning clássico, Aprendizado de representação e Deep Learning.



Tendências históricas em Deep Learning

Tendências Históricas em Deep Learning:

1. Deep learning tem uma rica história - teve vários nomes ao longo do tempo e diferentes de pontos de vista filosóficos;
2. Se tornou mais interessante quando aumentou-se a quantidade de dados de treinamento;
3. Hardware e software evoluíram, possibilitando os modelos a também evoluírem;
4. Deep Learning tem resolvido problemas mais complexos e vem melhorando a precisão ao longo do tempo.

Houveram três ondas:

- cibernética - entre 1940 até 1960
- connexionista - entre 1980 a 1990
- deep learning - de 2006 até hoje

- Redes neurais artificiais - inspiradas em um cérebro biológico.

- Redes neurais artificiais - inspiradas em um cérebro biológico.
- O cérebro provê uma prova de que um comportamento inteligente é possível.

- Redes neurais artificiais - inspiradas em um cérebro biológico.
- O cérebro provê uma prova de que um comportamento inteligente é possível.
- **1943** - McCulloch-Pitts descrevem o **neurônio** - os pesos eram setados manualmente

- Redes neurais artificiais - inspiradas em um cérebro biológico.
- O cérebro provê uma prova de que um comportamento inteligente é possível.
- **1943** - McCulloch-Pitts descrevem o **neurônio** - os pesos eram setados manualmente
- **1958-1962** - Roseblatt publica o **Perceptron** - podia aprender a ajustar os pesos de acordo com as entradas e saídas.

- Redes neurais artificiais - inspiradas em um cérebro biológico.
- O cérebro provê uma prova de que um comportamento inteligente é possível.
- **1943** - McCulloch-Pitts descrevem o **neurônio** - os pesos eram setados manualmente
- **1958-1962** - Roseblatt publica o **Perceptron** - podia aprender a ajustar os pesos de acordo com as entradas e saídas.
- **1960** - Widrow e Hoff publicam ADALINE para predição de números reais e o algoritmo de treinamento para ajustar os pesos do Adaline era um caso especial de um algoritmo chamado **gradiente descendente estocástico**.

- Tanto o Percetron com o Adaline são modelos **lineares**
- a limitação é que não aprendem o XOR (por ser não linear).

| A | B | A XOR B |
|---|---|----------------|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

- Tanto o Percetron com o Adaline são modelos **lineares**
- a limitação é que não aprendem o XOR (por ser não linear).
- Essa foi a primeira “grande queda” das redes neurais.

| A | B | A XOR B |
|---|---|----------------|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

Relação com a Neurociência

- A neurociência é uma importante fonte de inspiração para deep learning.



Relação com a Neurociência

- A neurociência é uma importante fonte de inspiração para deep learning.
- Porém está longe de ser o guia predominante.



Relação com a Neurociência

- A neurociência é uma importante fonte de inspiração para deep learning.
- Porém está longe de ser o guia predominante.
- Não temos informação suficiente sobre o cérebro e estamos longe de compreendê-lo e monitorar o que acontece nele.



Relação com a Neurociência

- Neurocientistas descobriram que **furões** podem aprender a ver com a região do processamento auditivo do cérebro se o cérebro for religado para enviar sinais visuais para esta área.



Relação com a Neurociência

- Neurocientistas descobriram que **furões** podem aprender a ver com a região do processamento auditivo do cérebro se o cérebro for religado para enviar sinais visuais para esta área.
- Isso sugere que o cérebro dos mamíferos pode usar um único algoritmo para resolver a maioria das diferentes tarefas que o cérebro resolve.



Relação com a Neurociência

- Neurocientistas descobriram que furões podem aprender a ver com a região do processamento auditivo do cérebro se o cérebro for religado para enviar sinais visuais para esta área.
- **Isso sugere que o cérebro dos mamíferos pode usar um único algoritmo para resolver a maioria das diferentes tarefas que o cérebro resolve.**
- Esta descoberta trouxe uma esperança e uma **reaproximação** dos diferentes grupos de IA - NLP, computational vision, speech recognition, etc.

- **Fukushima** trouxe duas contribuições muito importantes:
- Cognitron (1975): ideia do modelo de ativação de neurônio usado atualmente chamado de **rectified linear unit (ReLU)**.

Neurociência Computacional

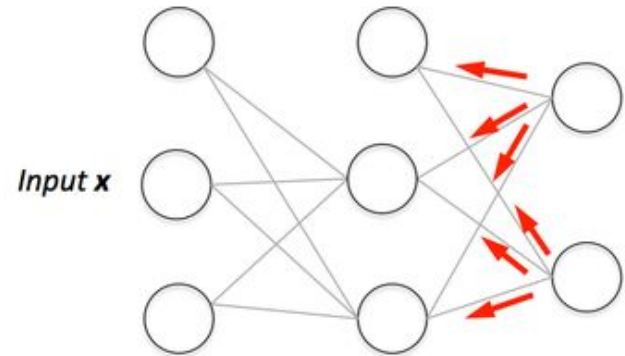
- Não podemos ver deep learning como uma tentativa de **simular o cérebro**.
- A área responsável por estudar como o cérebro funciona em um nível algorítmico é chamado de **neurociência computacional**.
- Deep Learning se preocupa em resolver tarefas que requerem inteligência, enquanto que neurociência computacional se preocupa primariamente com construir modelos cada vez mais precisos de como o cérebro realmente funciona.

Movimento Conexionista

- 1980 emerge o movimento conexionista.
- A ideia central é que um grande número de unidades computacionais simples podem atingir um comportamento inteligente quando conectadas entre si.
- Conceito de **Aprendizado Distribuído**: um neurônio aprende uma classe, ex: carrom caminhão e outro neurônio aprende a cor;

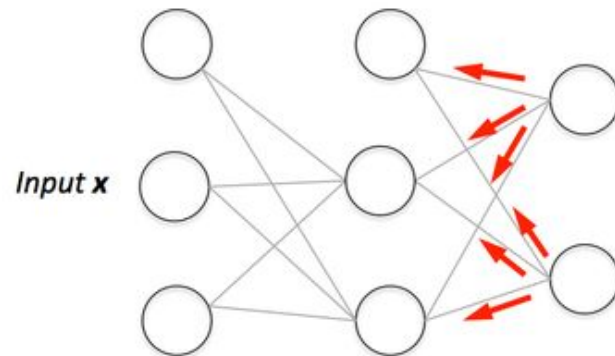
Movimento Conexcionista

- 1980 emerge o movimento conexionista.
- A ideia central é que um grande número de unidades computacionais simples podem atingir um comportamento inteligente quando conectadas entre si.
- Conceito de **Aprendizado Distribuído**: um neurônio aprende uma classe, ex: carrom caminhão e outro neurônio aprende a cor;
- Rumelhart (1986) e LeCun (1987) - **Back-propagation**



Movimento Conexcionista

- 1980 emerge o movimento conexionista.
- A ideia central é que um grande número de unidades computacionais simples podem atingir um comportamento inteligente quando conectadas entre si.
- Conceito de **Aprendizado Distribuído**: um neurônio aprende uma classe, ex: carrom caminhão e outro neurônio aprende a cor;
- Rumelhart (1986) e LeCun (1987) - **Back-propagation**
- 1997 LSTM foi proposta e hoje é usado para modelar
- tarefas que envolvem sequência, ex: NLP.



Os investidores desacreditaram das redes neurais, pois haviam propostas muito ambiciosas que não podiam ser atingidas.

Os investidores desacreditaram das redes neurais, pois haviam propostas muito ambiciosas que não podiam ser atingidas.

Muita promessa, pouco resultado.

Os investidores desacreditaram das redes neurais, pois haviam propostas muito ambiciosas que não podiam ser atingidas.

Muita promessa, pouco resultado.

Além disso, as redes profundas eram consideradas muito difíceis de treinar, visto a quantidade de dados e poder computacional da época.

Terceira Onda

Popularizou o termo deep learning para enfatizar que os pesquisadores estavam interessados agora em focar a atenção na importância teórica da profundidade e que agora tinham acesso a treinar as redes neurais profundas.

Terceira Onda

Popularizou o termo deep learning para enfatizar que os pesquisadores estavam interessados agora em focar a atenção na importância teórica da profundidade e que agora tinham acesso a treinar as redes neurais profundas.

As redes neurais profundas começaram a superar outros métodos de IA em competições.

Terceira Onda

- Conforme a sociedade se digitaliza temos mais dados - **Era do Big Data**.
- Habilidade necessária diminuiu conforme o aumento do tamanho dos **datasets**.



Terceira Onda

- Aumento no tamanho dos modelos pois hoje temos recurso computacional para rodar modelos maiores.
- GPUs foram uma evolução e computação distribuída são tendências.



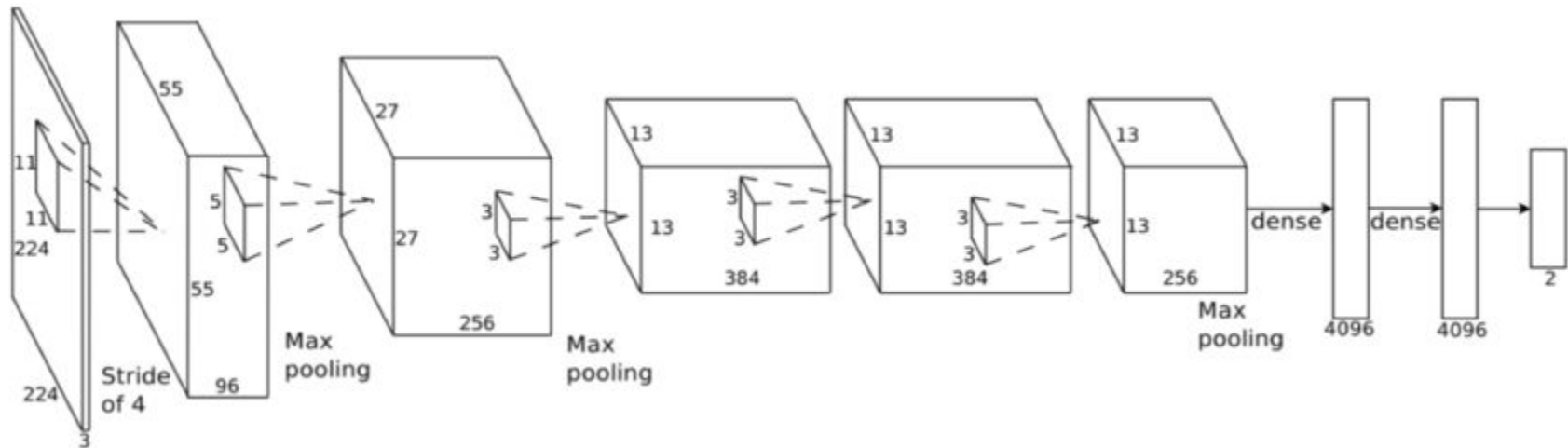
Terceira Onda

- A maior competição de object detection é a ImageNet Large Scale Visual Recognition Challenge .



Terceira Onda

- A maior competição de object detection é a ImageNet Large Scale Visual Recognition Challenge .
- O momento histórico foi quando a **AlexNet**, uma rede convolucional, ganhou o desafio, baixando a taxa de erro de 26,1% para 15,3% em 2012.

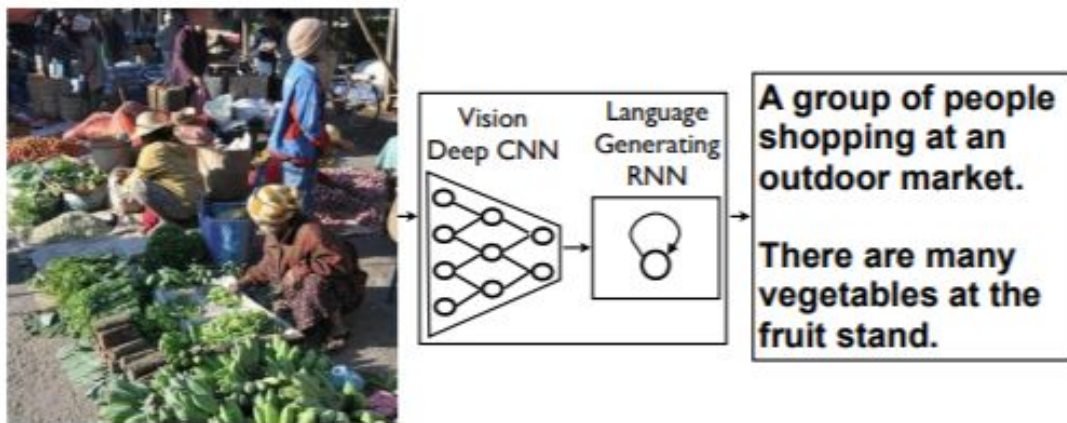


- A maior competição de object detection é a ImageNet Large Scale Visual Recognition Challenge .
- O momento histórico foi quando a **AlexNet**, uma rede convolucional, ganhou o desafio, baixando a taxa de erro de 26,1% para 15,3% em 2012.
- Desde então, só redes profundas ganham a competição e hoje a taxa de erro está em 3,6%.

- A comunidade de speech recognition também adotou redes profundas e a taxa de erro caiu mais do que pela metade.

Terceira Onda

- A comunidade de speech recognition também adotou redes profundas e a taxa de erro caiu mais do que pela metade.
- Aumentou-se também a complexidade das tarefas - ex: show and tell, machine translation, neural turing machines, etc.



Terceira Onda

- **Aprendizado por Reforço:** DeepMind mostrou que é capaz de jogar vídeo game atingindo performance de humanos.



Terceira Onda

- **Aprendizado por Reforço:** DeepMind mostrou que é capaz de jogar vídeo game atingindo performance de humanos.



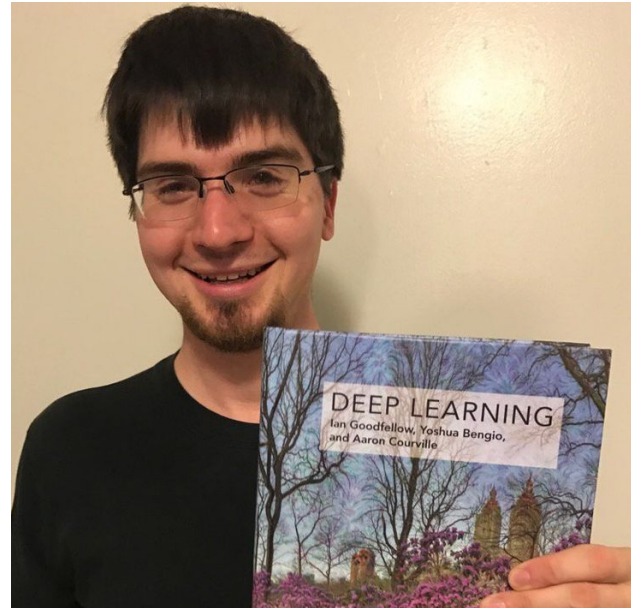
Terceira Onda

- **Aprendizado por Reforço:** DeepMind mostrou que é capaz de jogar vídeo game atingindo performance de humanos.



- **Aprendizado por Reforço:** DeepMind mostrou que é capaz de jogar vídeo game atingindo performance de humanos.
- Deep Learning também provê ferramentas para outras áreas, como por exemplo como as moléculas vão interagir a fim de ajudar farmacêuticos a desenvolverem novas drogas, busca por partículas subatômicas, etc.

“The years ahead are full of challenges and opportunities to improve deep learning even further and bring it to new frontiers.” Goodfellow, Ian



<http://abre.ai/grupo-prisma-capitulo-1>