

Redes Neurais e Aprendizagem Profunda

BREVE HISTÓRICO

Zenilton K. G. Patrocínio Jr

zenilton@pucminas.br

Inteligência Artificial

Um “sonho” antigo

- Criar máquinas capazes de pensar

Inteligência Artificial

Um “sonho” antigo

- Criar máquinas capazes de pensar

Abordagem IA Tradicional (“primórdios”)

- Um problema é descrito por um conjunto de regras matemáticas formais
- IA rapidamente abordou e resolveu problemas difíceis para os seres humanos
- Mas falhou em tarefas fáceis para a maioria das pessoas, como reconhecer palavras faladas ou rostos em imagens

Inteligência Artificial

Um “sonho” antigo

- Criar máquinas capazes de pensar

Possível solução

- Permitir que os computadores aprendam com a experiência e entendam o mundo em termos de uma hierarquia de conceitos
- Conceitos complexos (em níveis mais altos) são definidos em termos de sua relação com conceitos mais simples (em níveis mais baixos)

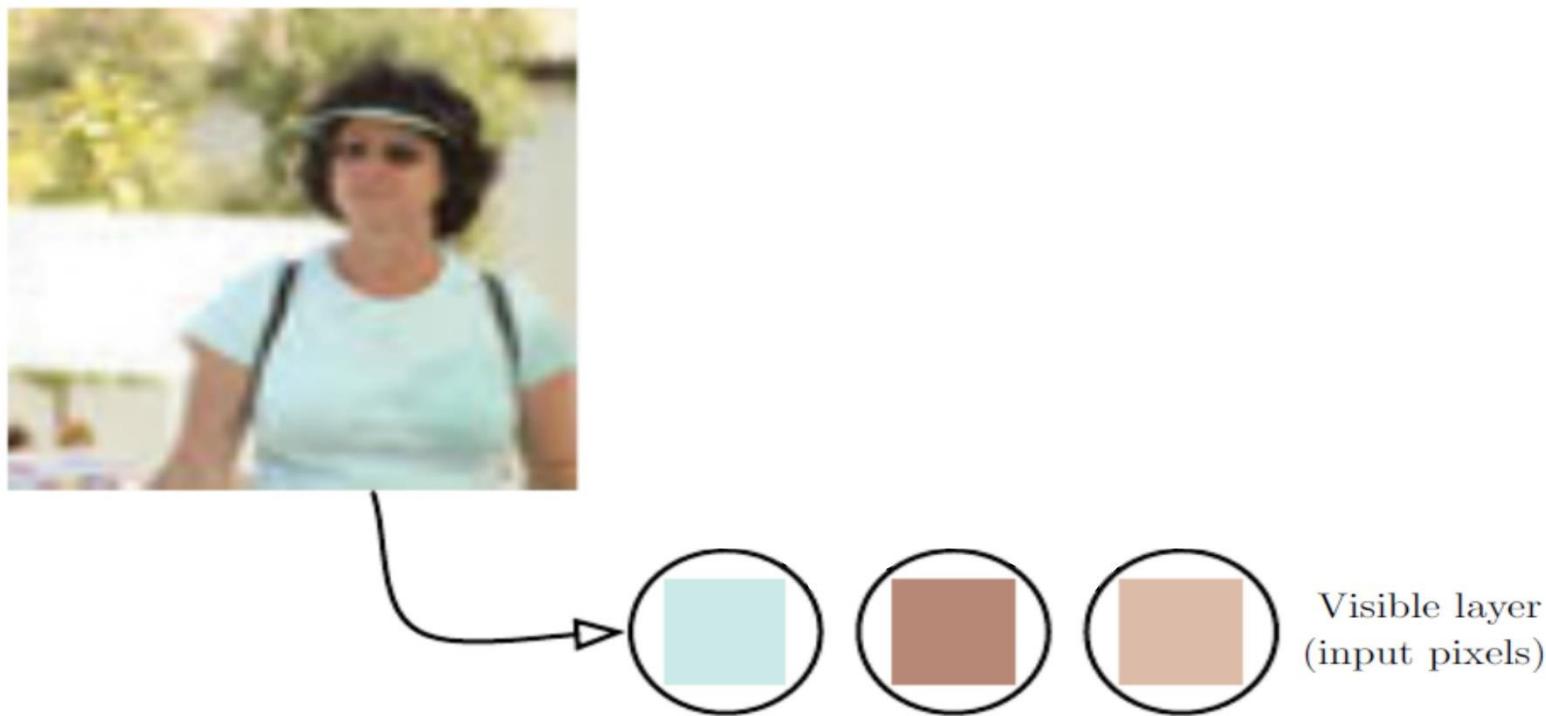
Um grafo mostrando como os conceitos estão uns sobre os outros será profundo

⇒ DEEP LEARNING

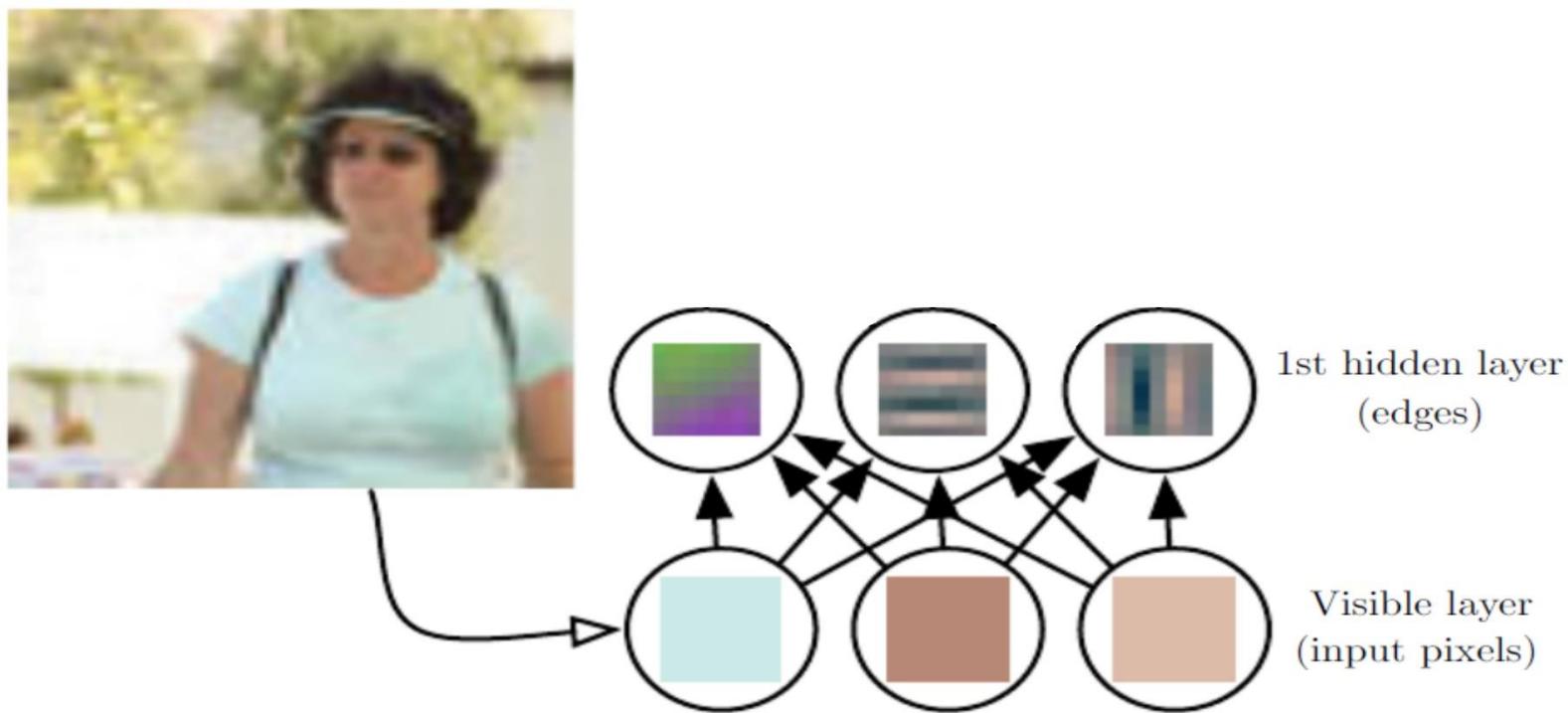
Aprendizagem Profunda como Aprendizagem de Representação



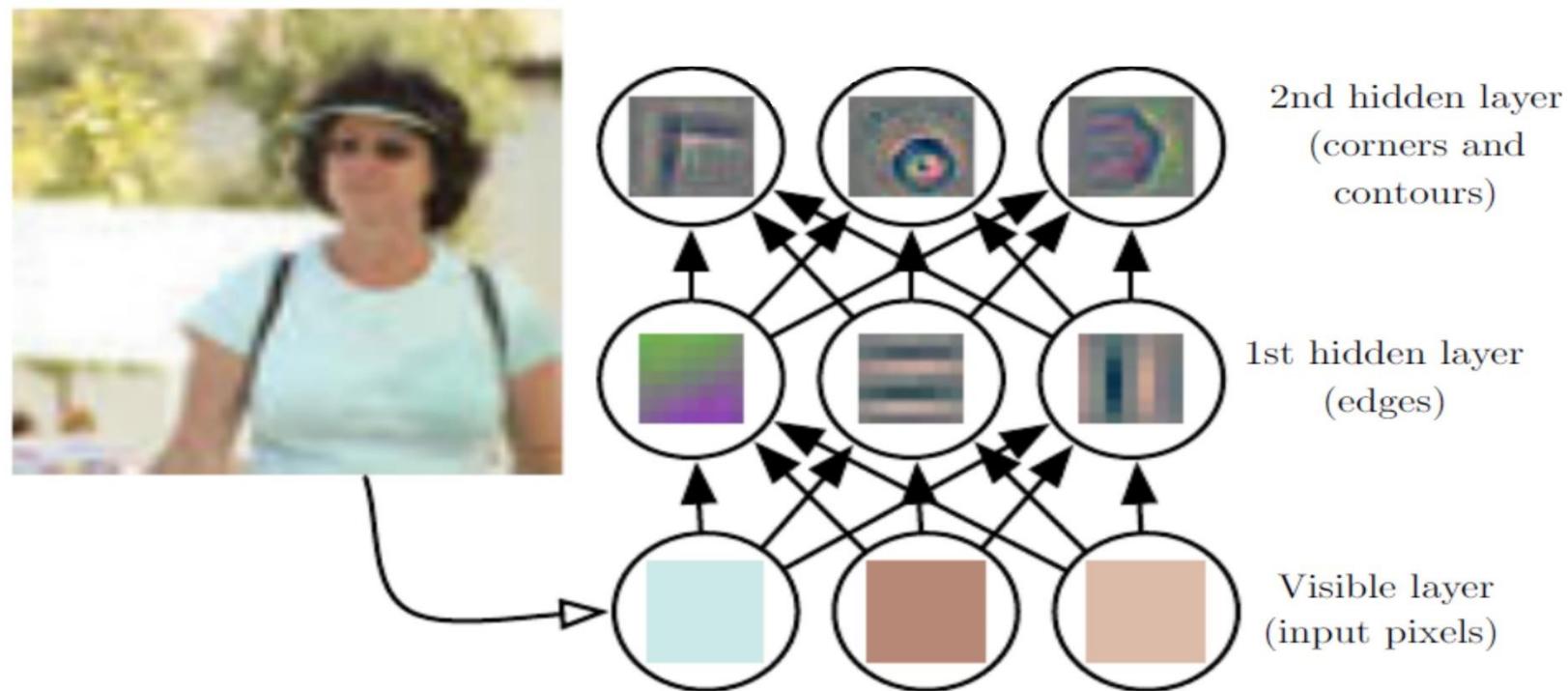
Aprendizagem Profunda como Aprendizagem de Representação



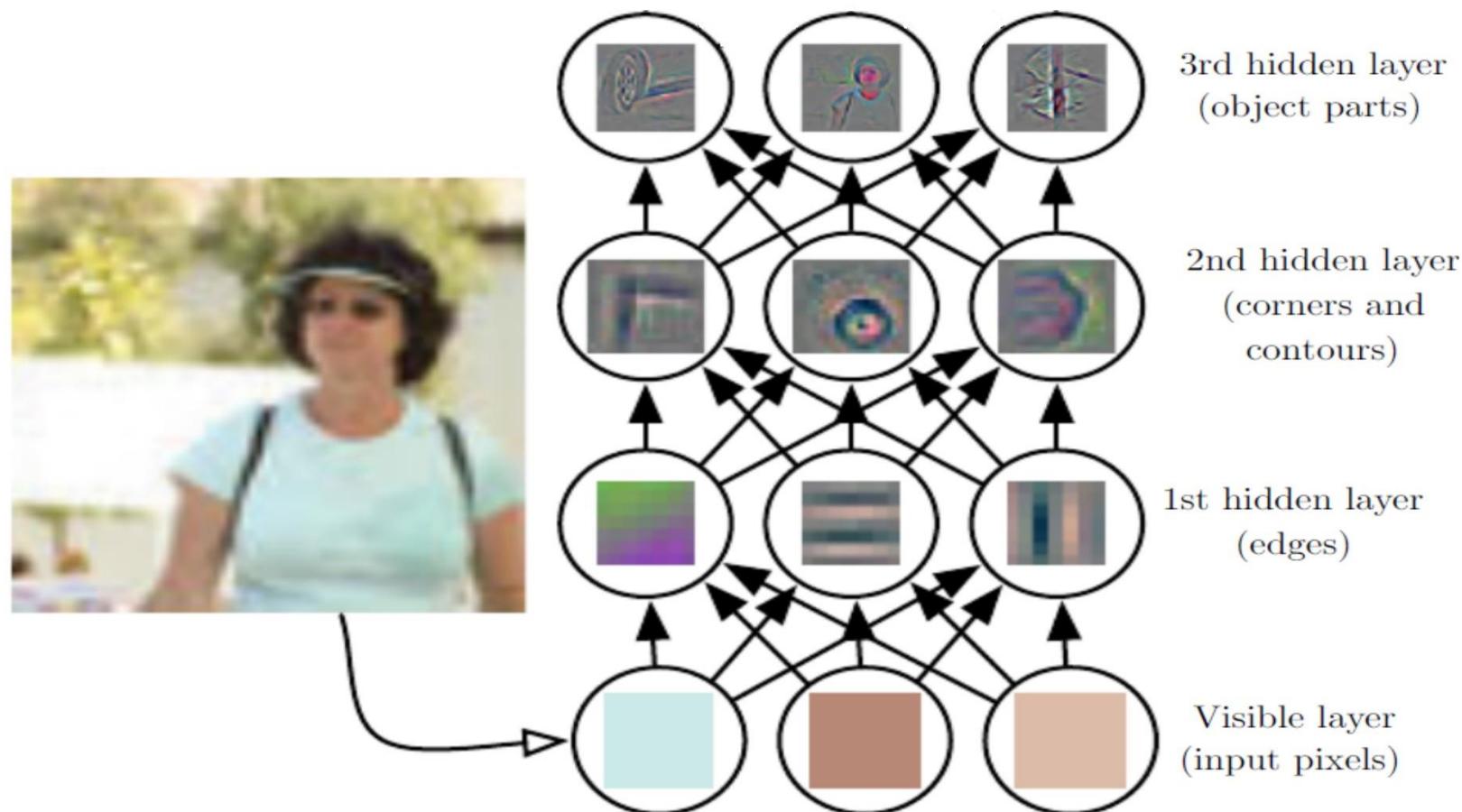
Aprendizagem Profunda como Aprendizagem de Representação



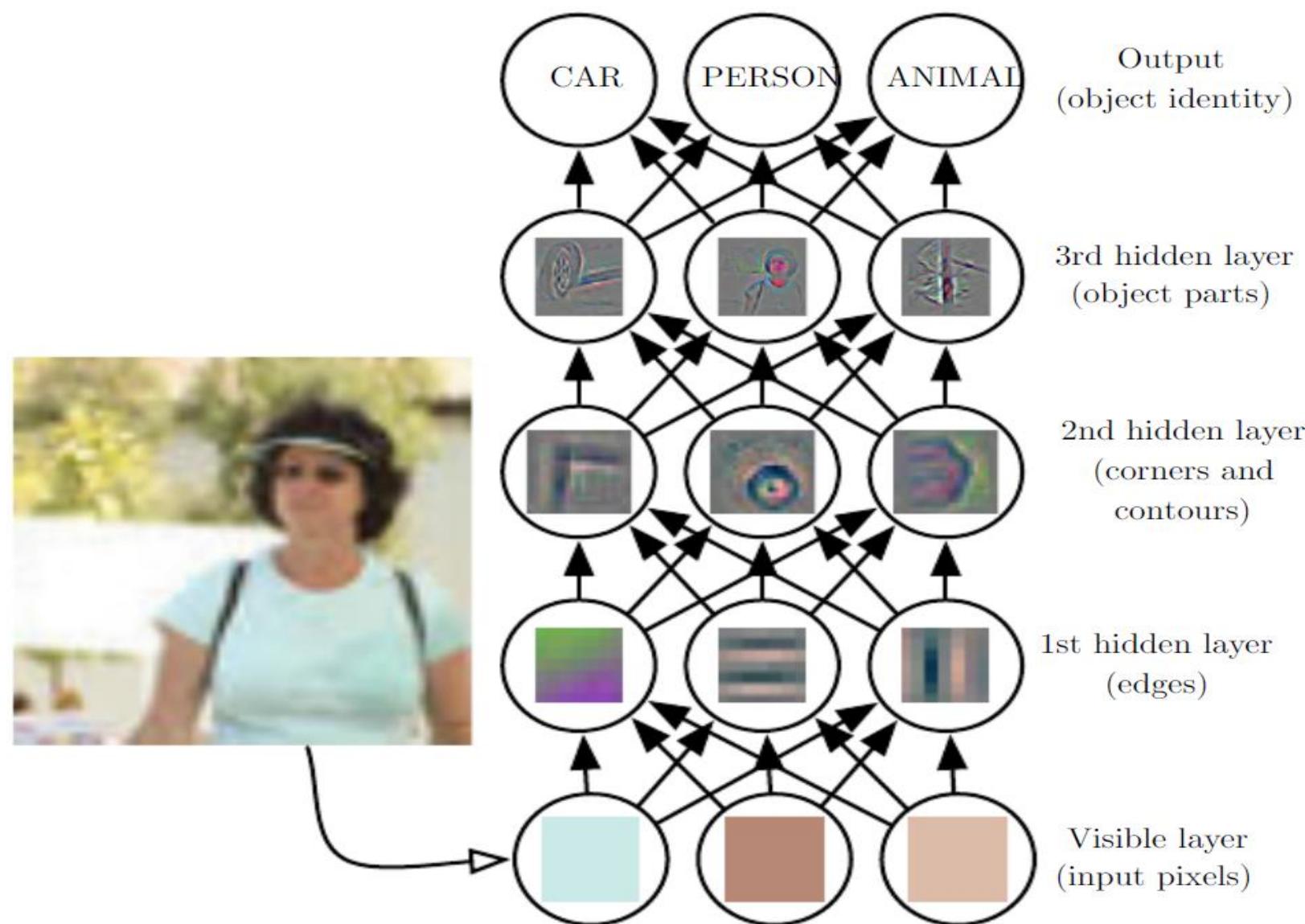
Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação



Aprendizado de Máquina

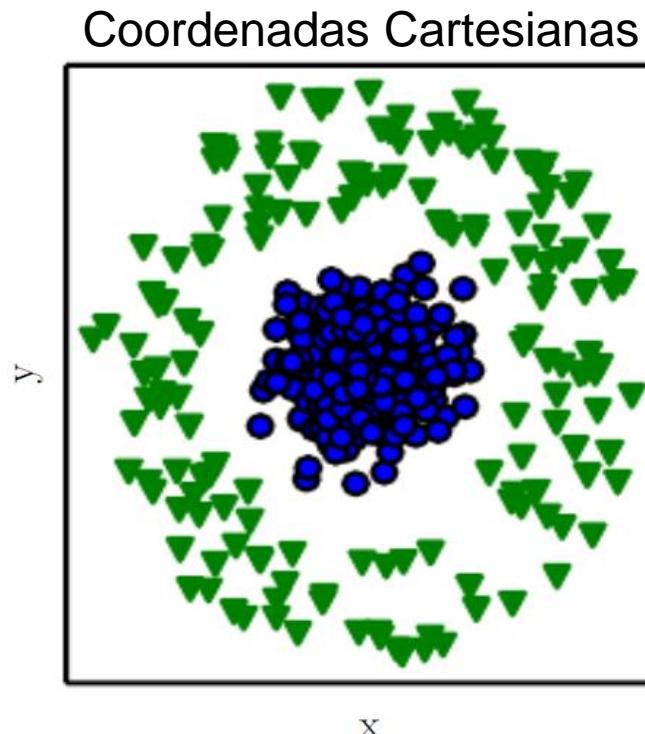
Aquisição de conhecimento a partir dos dados

- Extrair padrão dos dados brutos (em vez de uma solução codificada por regras)
- Desempenho depende muito da representação de dados

Aprendizado de Máquina

Aquisição de conhecimento a partir dos dados

- Extrair padrão dos dados brutos (em vez de uma solução codificada por regras)
- Desempenho depende muito da representação de dados

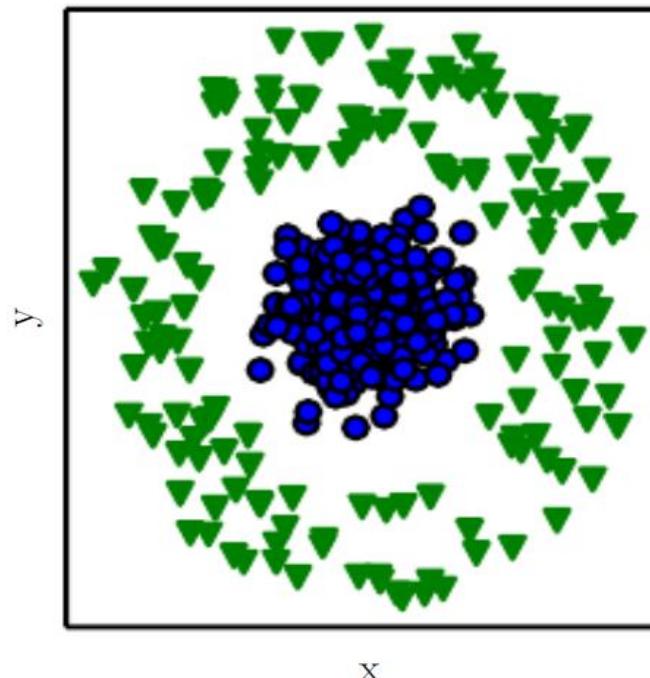


Aprendizado de Máquina

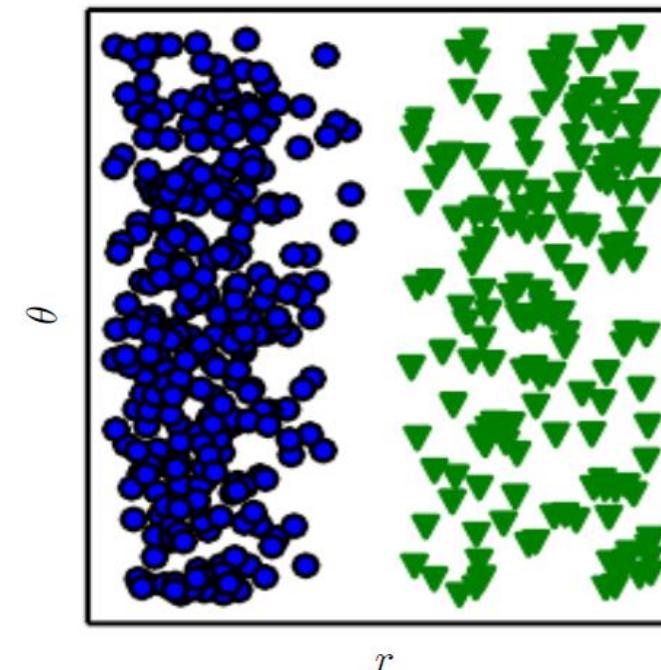
Aquisição de conhecimento a partir dos dados

- Extrair padrão dos dados brutos (em vez de uma solução codificada por regras)
- Desempenho depende muito da representação de dados

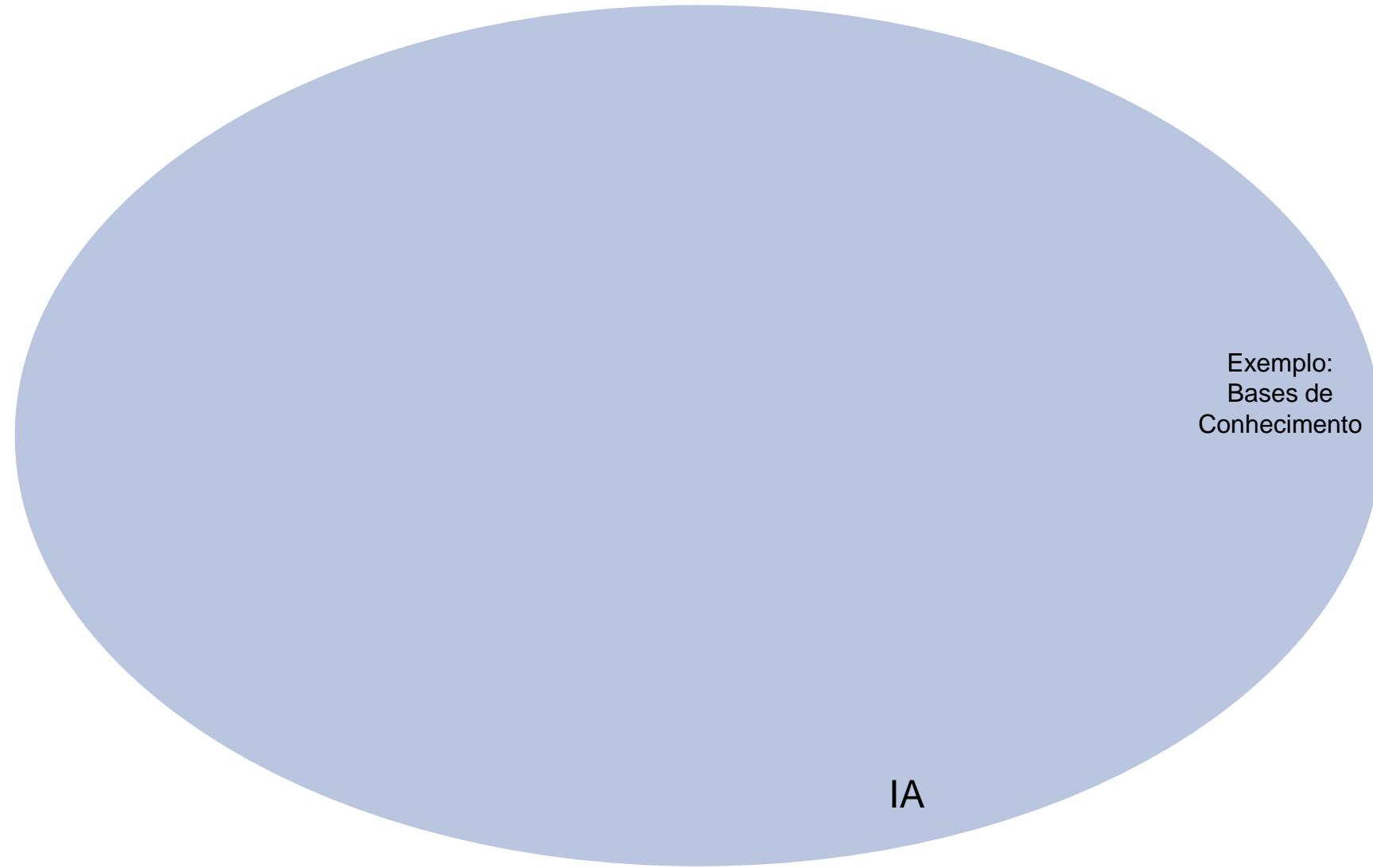
Coordenadas Cartesianas



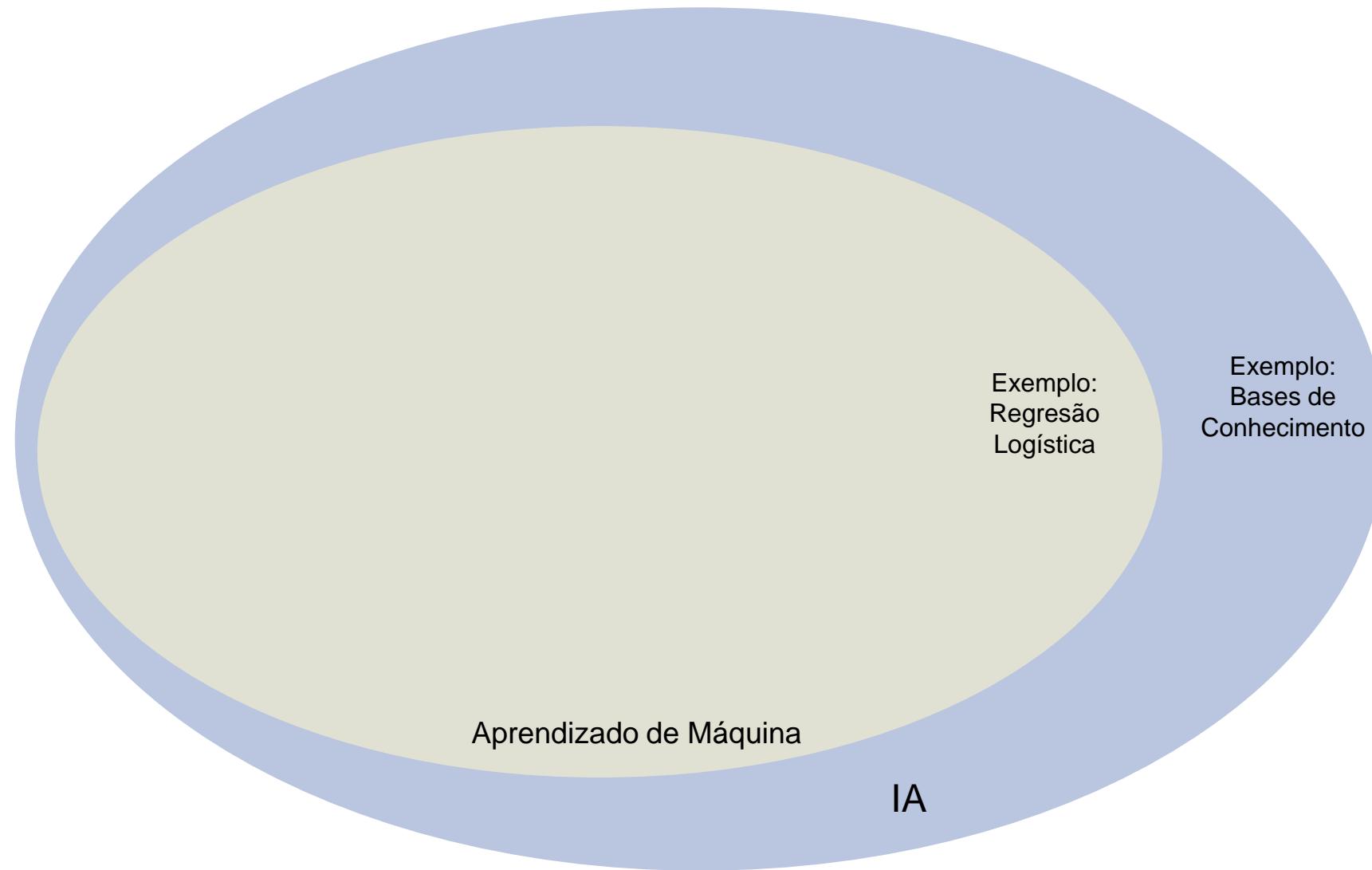
Coordenadas Polares



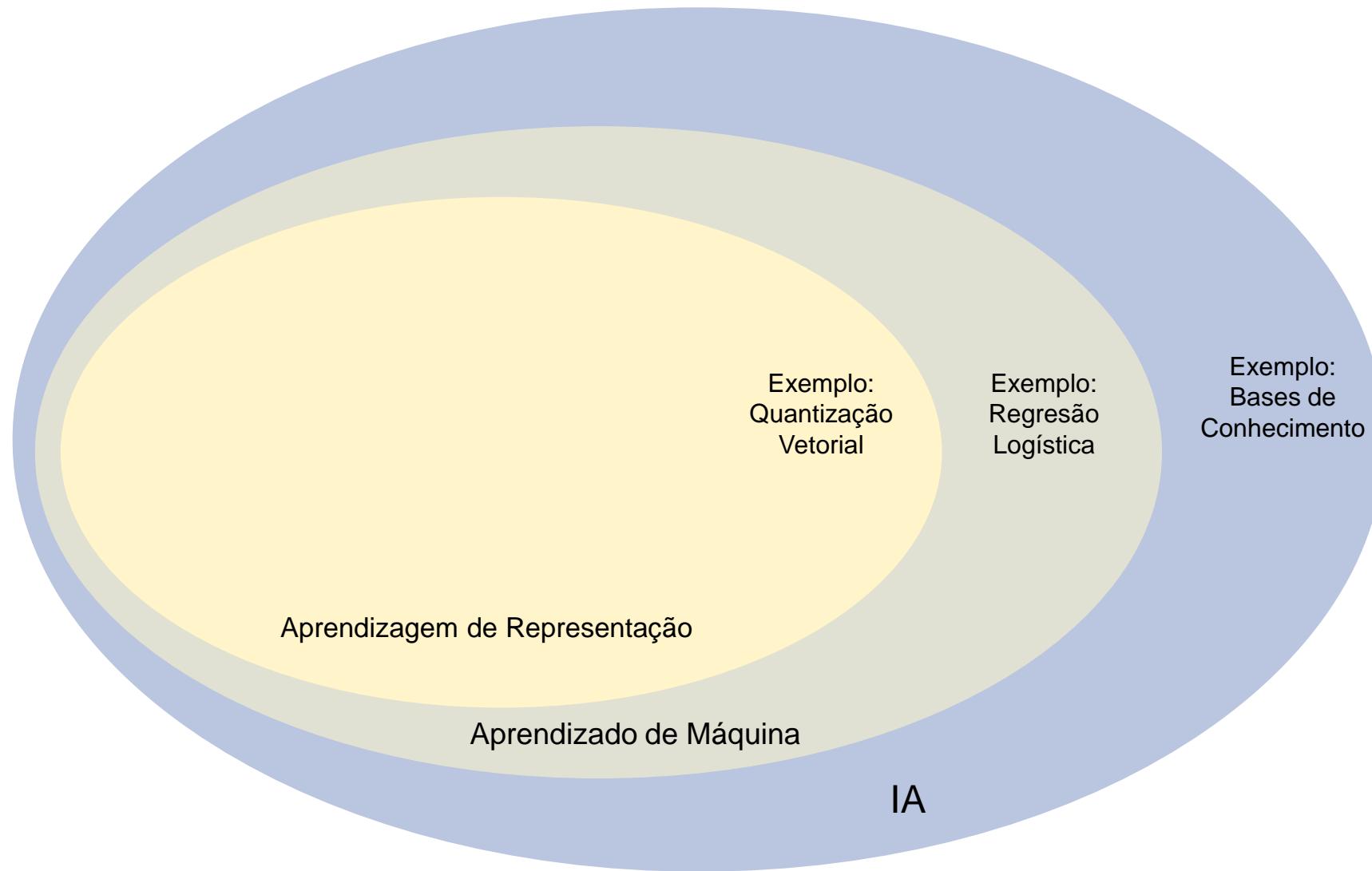
Aprendizagem Profunda como Aprendizagem de Representação



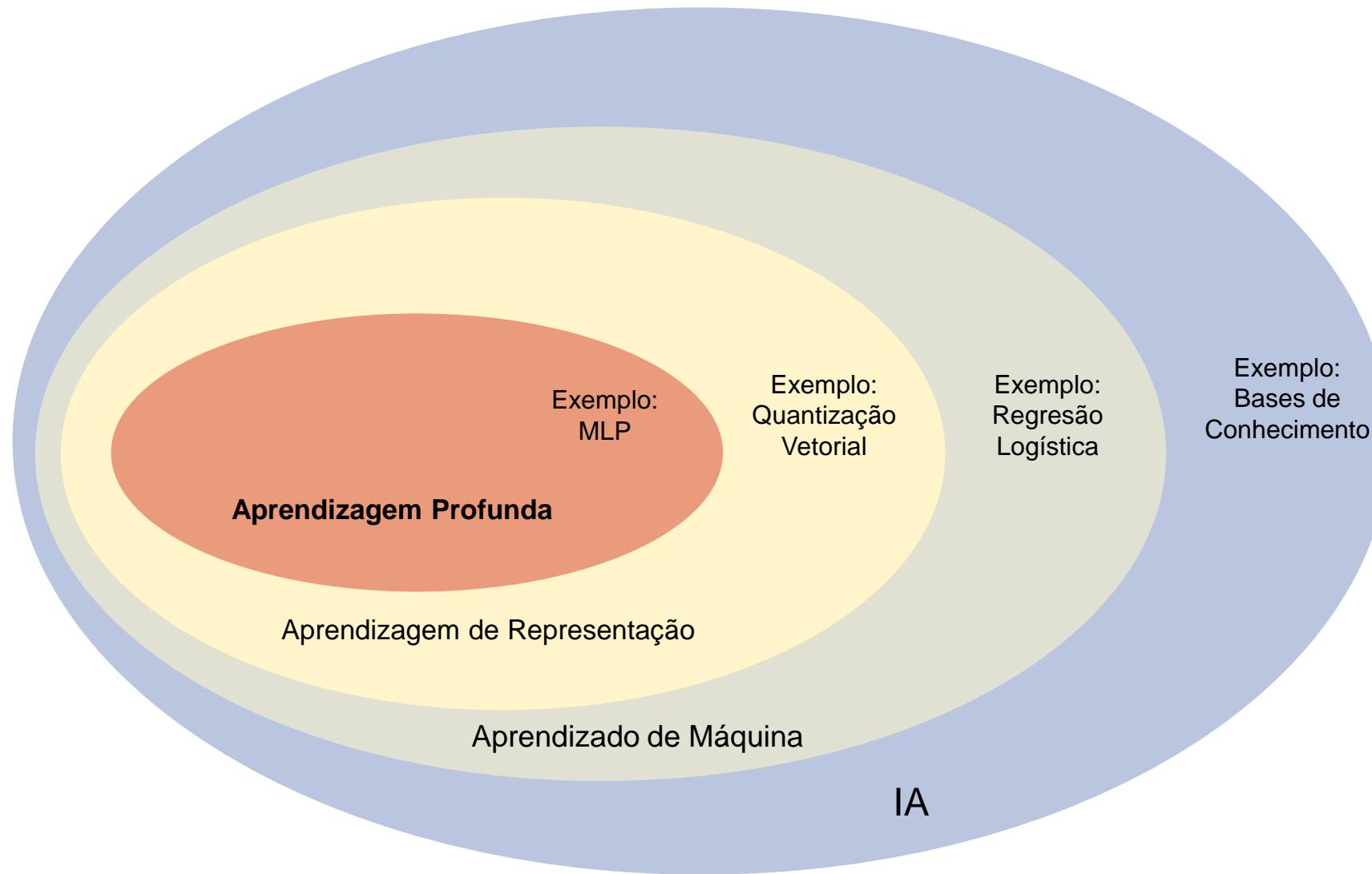
Aprendizagem Profunda como Aprendizagem de Representação



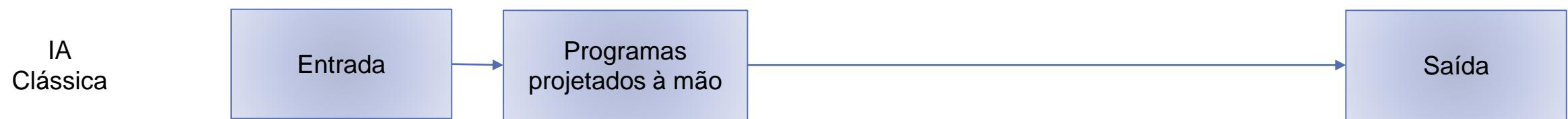
Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação

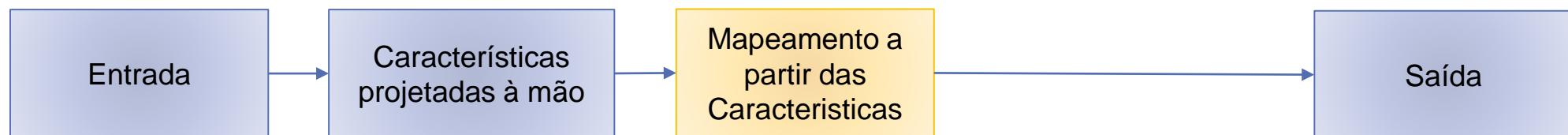


Aprendizagem Profunda como Aprendizagem de Representação

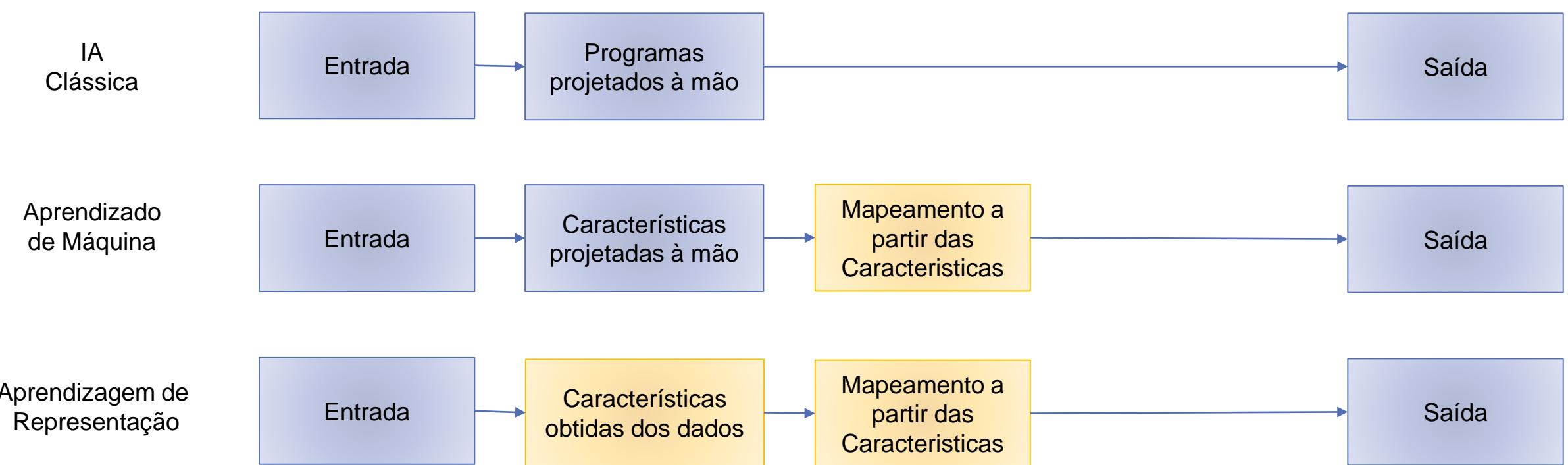
IA
Clássica



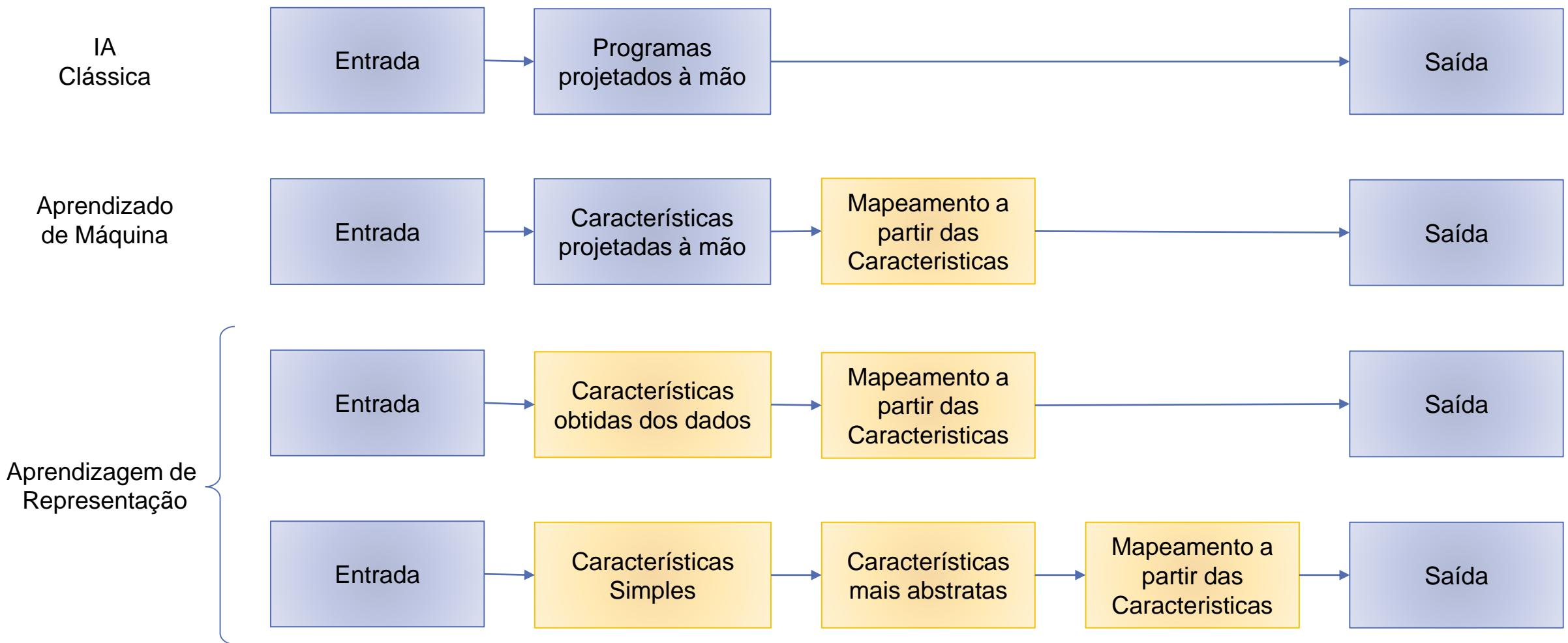
Aprendizado
de Máquina



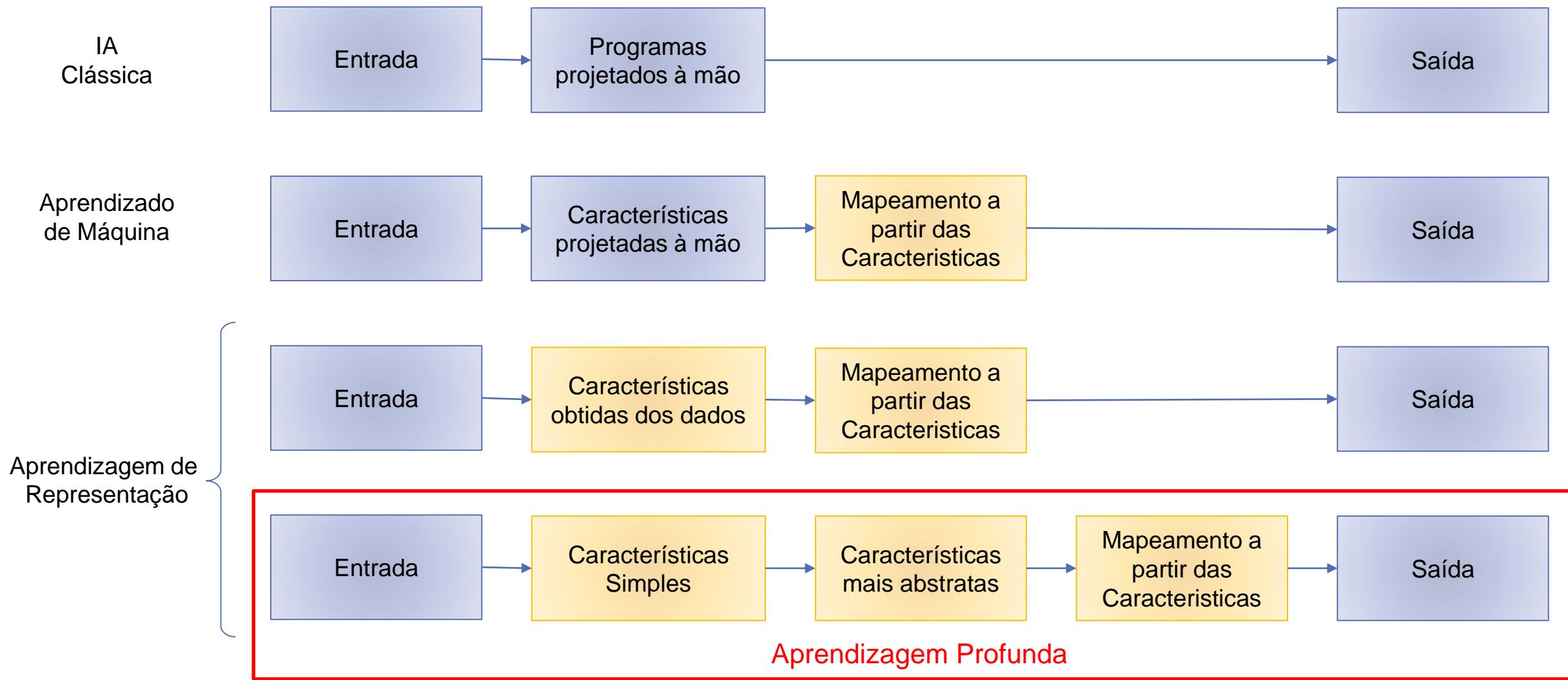
Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação



Aprendizagem Profunda como Aprendizagem de Representação



Fases da Pesquisa em Redes Neurais

- 1940s-1960s: **Cibernética**. época do “Cérebro eletrônico”, deu origem a moderna teoria de controle e processamento de sinais

Fases da Pesquisa em Redes Neurais

- 1940s-1960s: **Cibernética**. época do “Cérebro eletrônico”, deu origem a moderna teoria de controle e processamento de sinais
- 1960s-1980s: Computadores digitais, teoria de autômatos e teoria de complexidade computacional. **Simples circuitos ‘rasos’ apresentam grandes limitações na capacidade de representação**

Fases da Pesquisa em Redes Neurais

- 1940s-1960s: **Cibernética**. época do “Cérebro eletrônico”, deu origem a moderna teoria de controle e processamento de sinais
- 1960s-1980s: Computadores digitais, teoria de autômatos e teoria de complexidade computacional. **Simples circuitos ‘rasos’ apresentam grandes limitações na capacidade de representação**
- 1980s-1990s: **Conexionismo**. Redes complexas e não-lineares, advento da propagação retrograda (*back-propagation*)

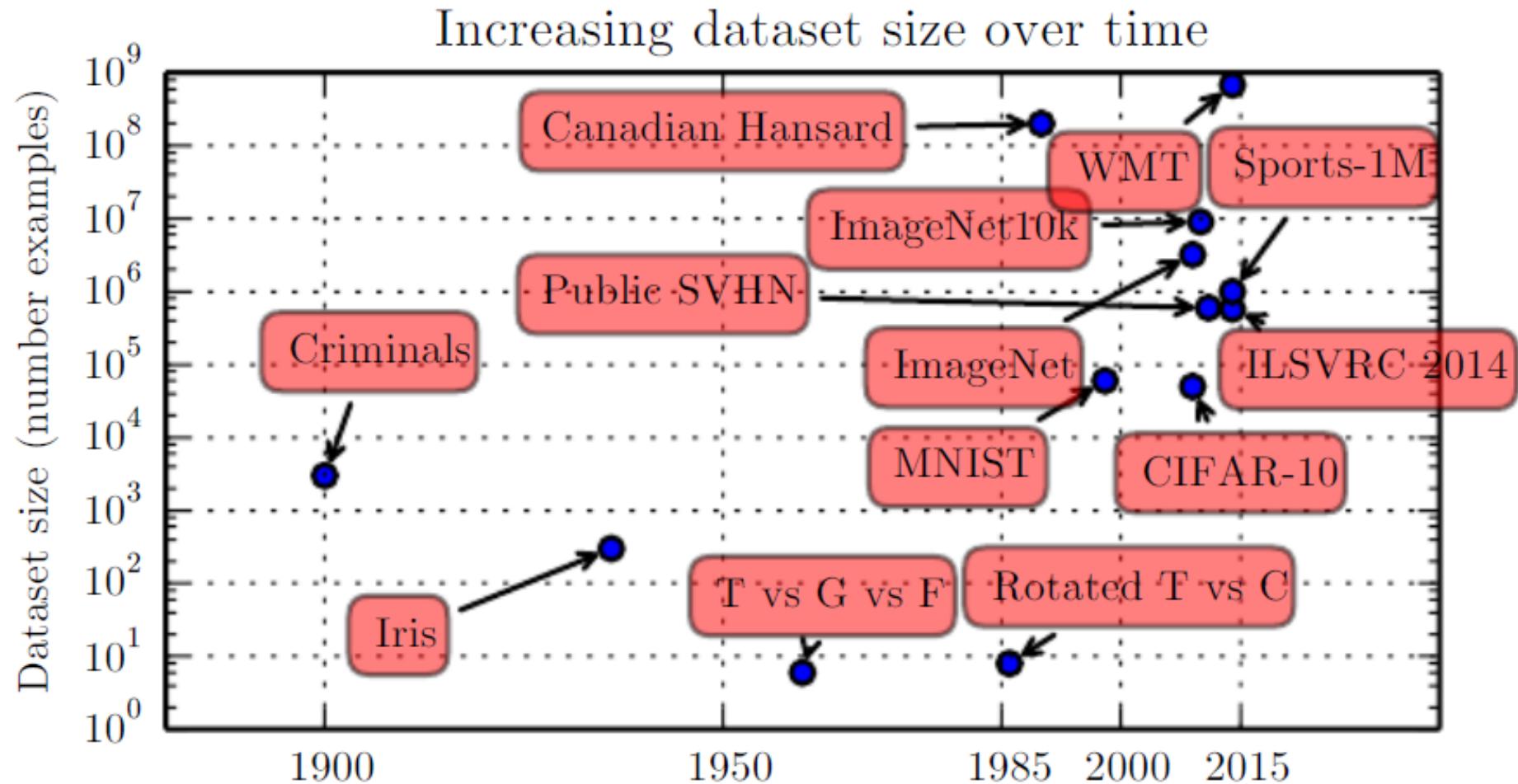
Fases da Pesquisa em Redes Neurais

- 1990s-2010s: Teoria de aprendizagem computacional, modelos gráficos.
Aprendizagem é um problema computacionalmente difícil, simples circuitos
'rasos' apresentam grandes limitações na capacidade de representação

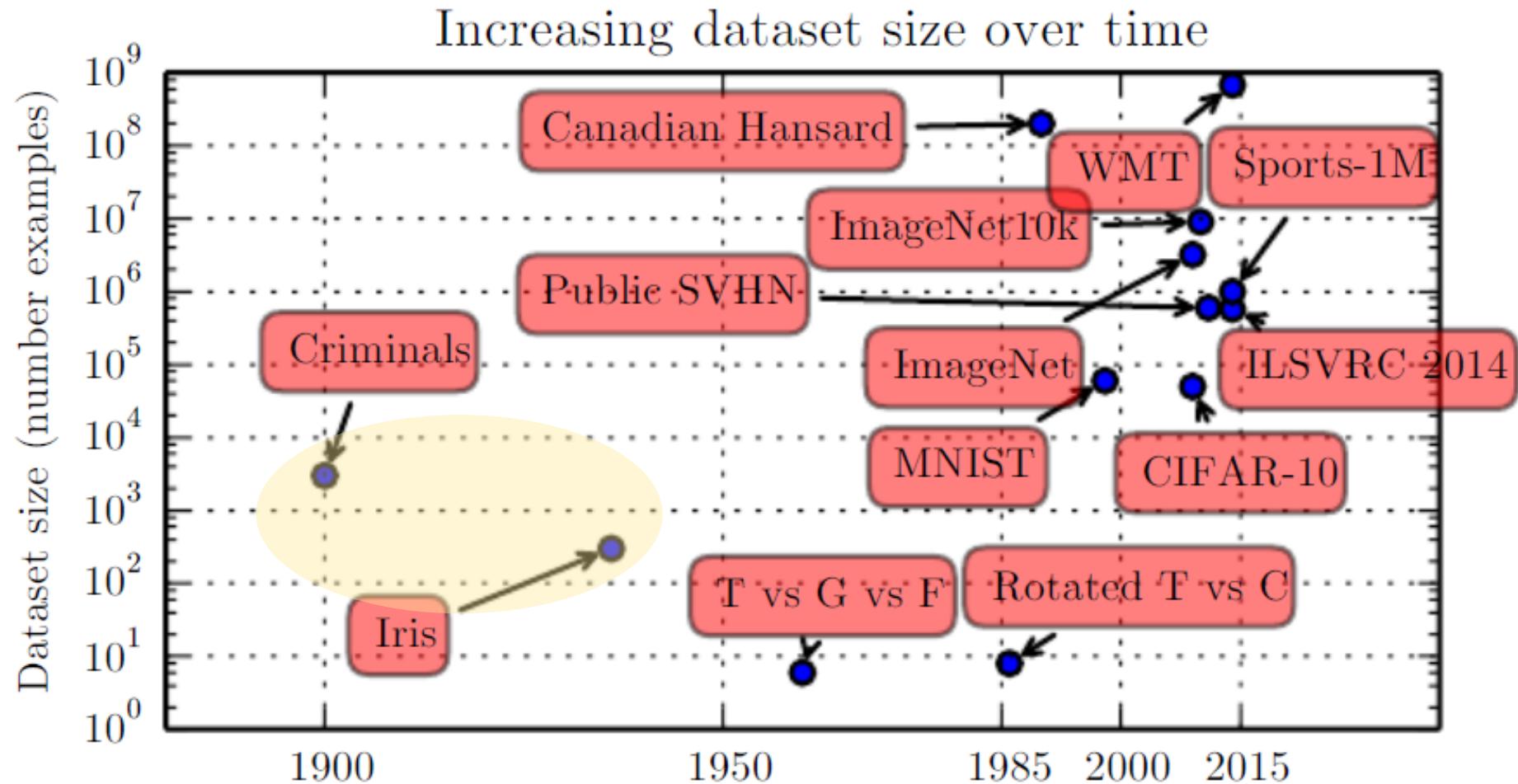
Fases da Pesquisa em Redes Neurais

- 1990s-2010s: Teoria de aprendizagem computacional, modelos gráficos.
Aprendizagem é um problema computacionalmente difícil, simples circuitos ‘rasos’ apresentam grandes limitações na capacidade de representação
- 2006→ . . . : Aprendizagem Profunda. Treinamento de redes neurais profundas ponta-a-ponta, grandes conjuntos de dados, explosão das aplicações

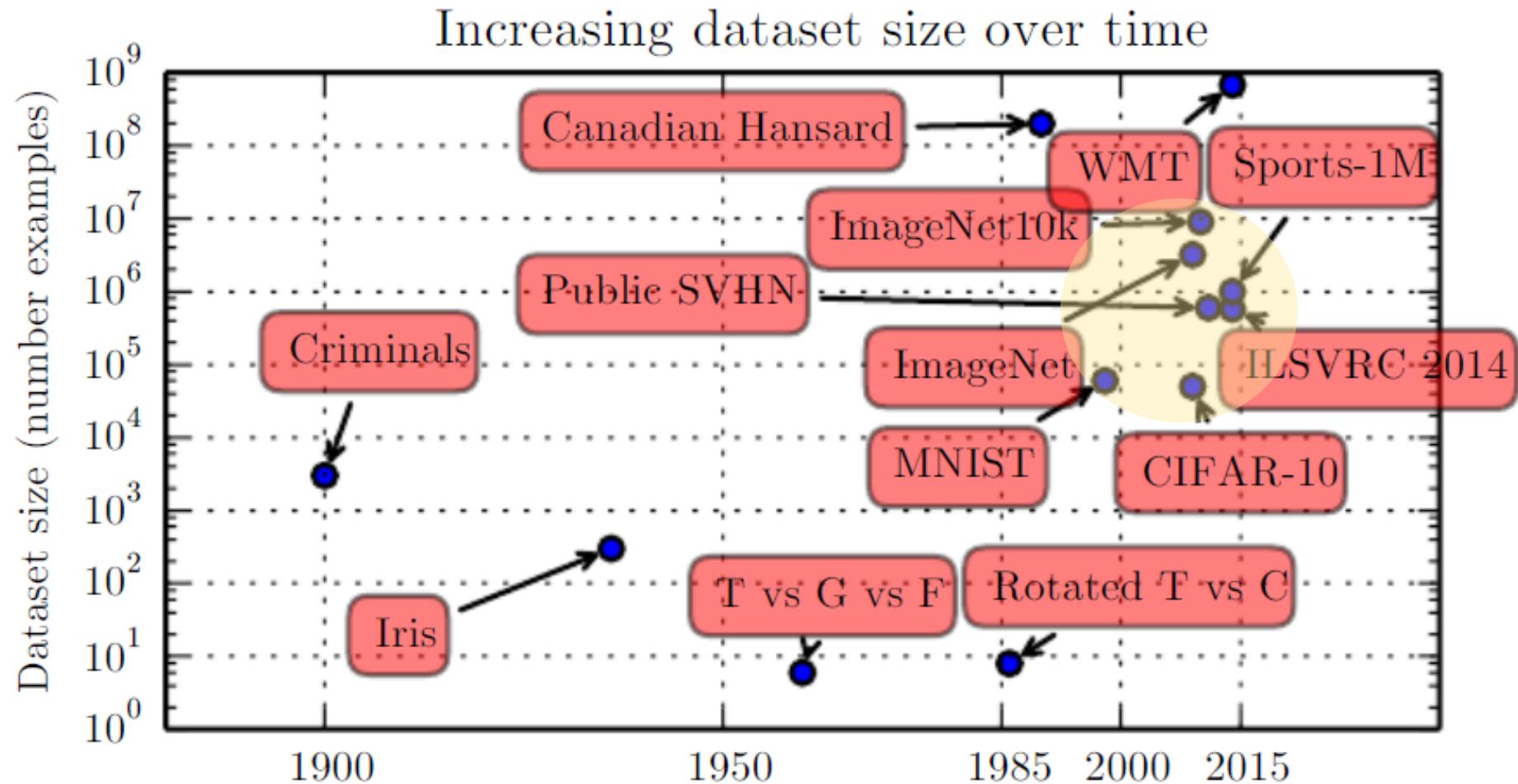
Razões para Sucesso: Aumento dos Conjuntos de Dados



Razões para Sucesso: Aumento dos Conjuntos de Dados



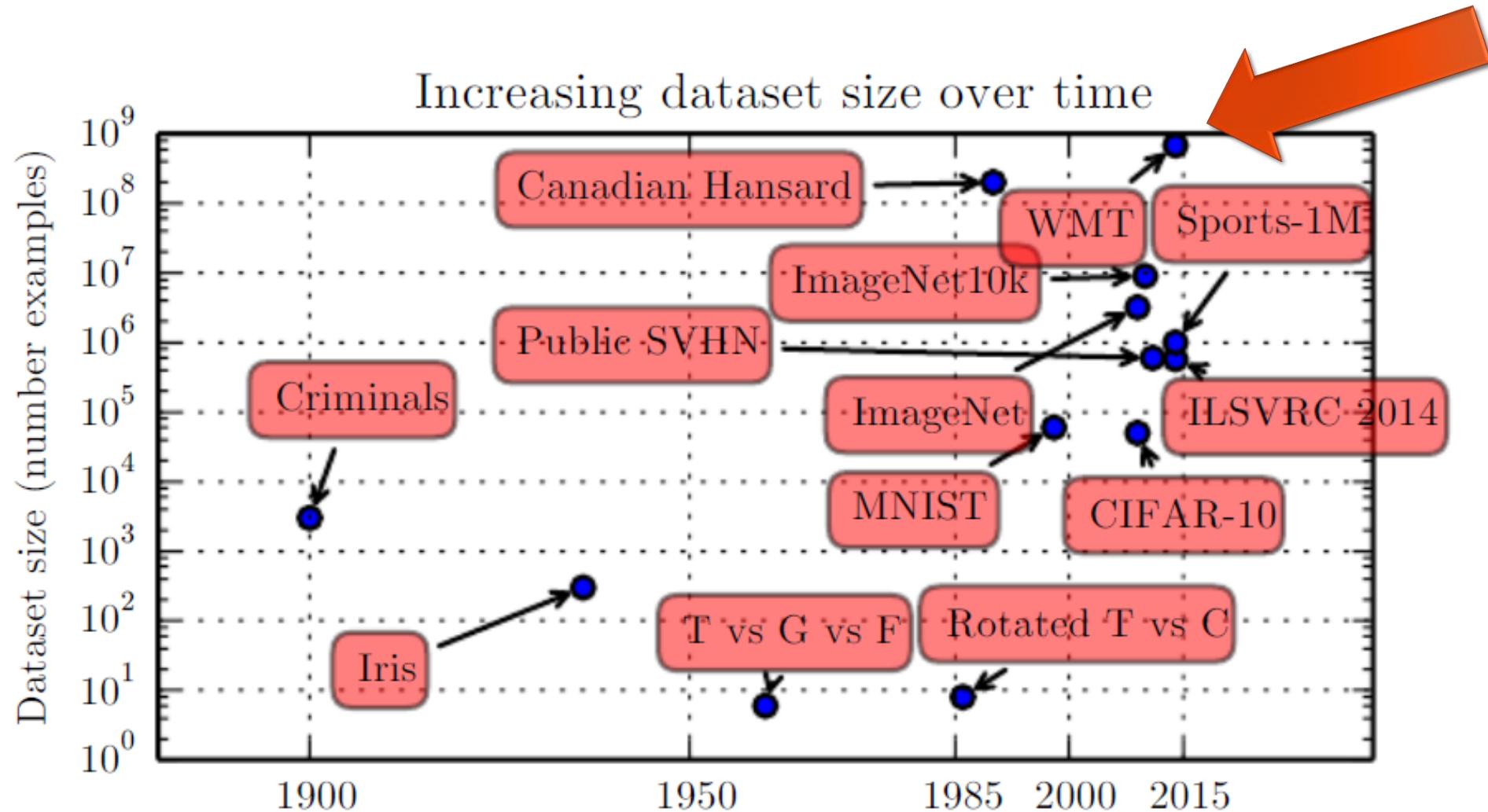
Razões para Sucesso: Aumento dos Conjuntos de Dados



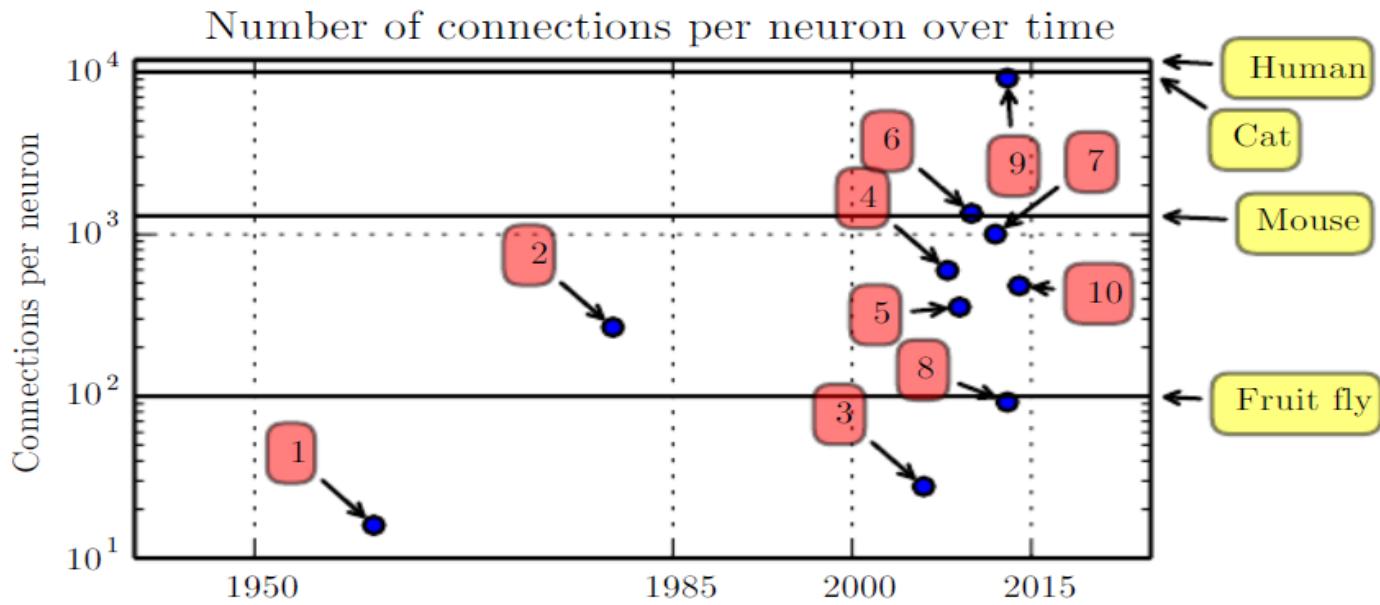
Exemplo: MNIST – “drosophila” do Aprendizado de Máquina

8	9	0	1	2	3	4	7	8	9	0	1	2	3	4	5	6	7	8	6
4	2	6	4	7	5	5	4	7	8	9	2	9	3	9	3	8	2	0	5
0	1	0	4	2	6	5	3	5	3	8	0	0	3	4	1	5	3	0	8
3	0	6	2	7	1	1	8	1	7	1	3	8	9	7	6	7	4	1	6
7	5	1	7	1	9	8	0	6	9	4	9	9	3	7	1	9	2	2	5
3	7	8	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	0
1	2	3	4	5	6	7	8	9	8	1	0	5	5	1	9	0	4	1	9
3	8	4	7	7	8	5	0	6	5	5	3	3	3	9	8	1	4	0	6
1	0	0	6	2	1	1	3	2	8	8	7	8	4	6	0	2	0	3	6
8	7	1	5	9	9	3	2	4	9	4	6	5	3	2	8	5	5	9	4
6	5	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8	9	6	4	2	6	4	7	5	5
4	7	8	9	2	9	3	9	3	8	2	0	9	8	0	5	6	0	1	0
4	2	6	5	5	5	4	3	4	1	5	3	0	8	3	0	6	2	7	1
1	8	1	7	1	3	8	5	4	2	0	9	7	6	7	4	1	6	8	4
7	5	1	2	6	7	1	9	8	0	6	9	4	9	9	6	2	3	7	1
9	2	2	5	3	7	8	0	1	2	3	4	5	6	7	8	0	1	2	3
4	5	6	7	8	0	1	2	3	4	5	6	7	8	9	2	1	2	1	3
9	9	8	5	3	7	0	7	7	5	7	9	9	4	7	0	3	4	1	4
4	7	5	8	1	4	8	4	1	8	6	6	4	6	3	5	7	2	5	9

Razões para Sucesso: Aumento dos Conjuntos de Dados

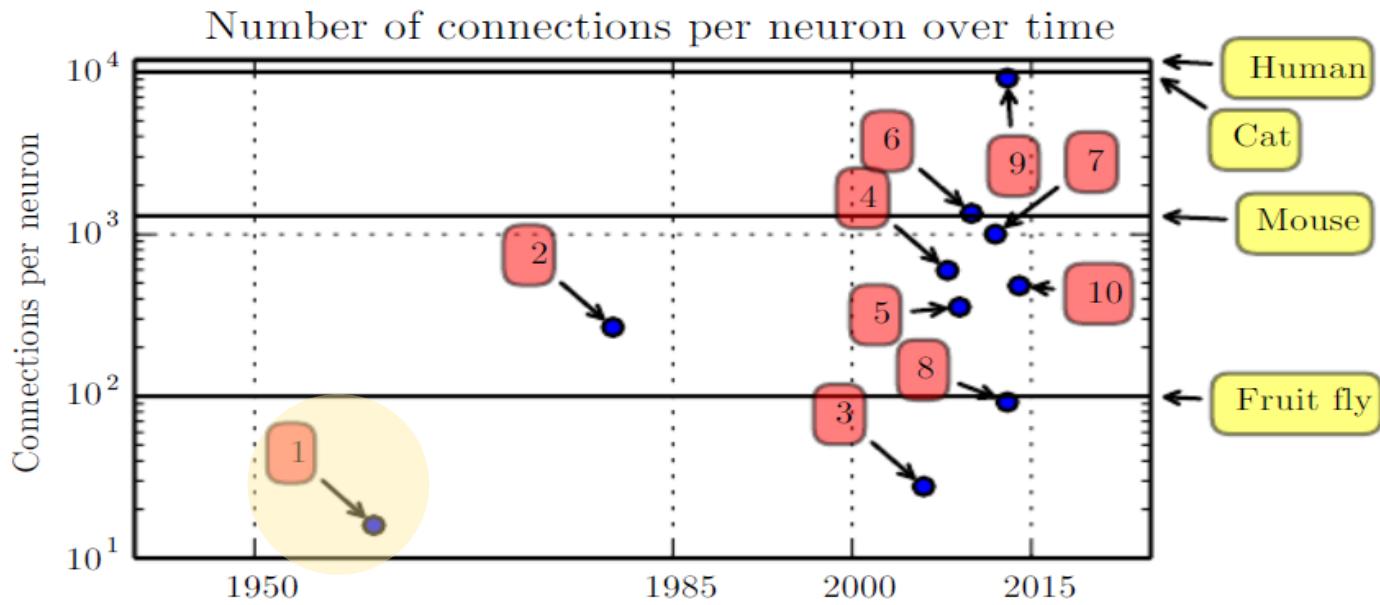


Razões para Sucesso: Aumento do Tamanho de Modelos



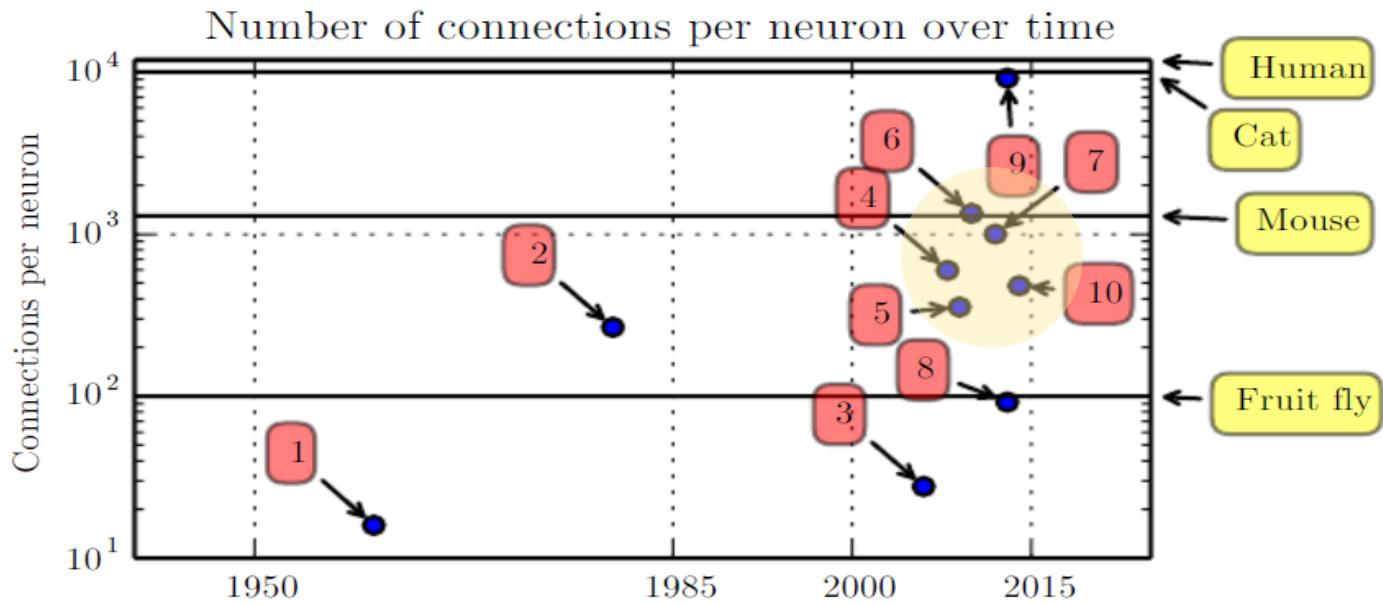
1. Adaptive linear element (Widrow and Hoff, 1960)
2. Neocognitron (Fukushima, 1980)
3. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
4. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
5. Unsupervised convolutional network (Jarrett *et al.*, 2009)
6. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
7. Distributed autoencoder (Le *et al.*, 2012)
8. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
9. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
10. GoogLeNet (Szegedy *et al.*, 2014a)

Razões para Sucesso: Aumento do Tamanho de Modelos



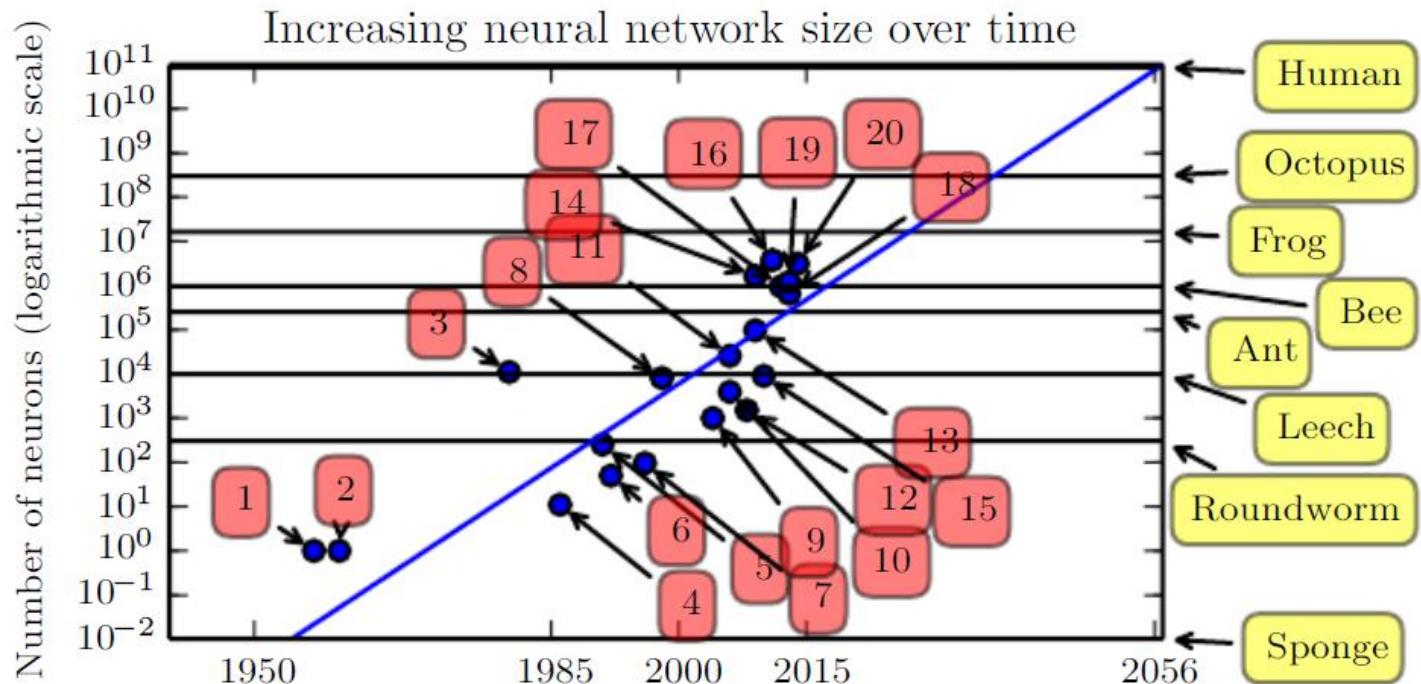
1. Adaptive linear element (Widrow and Hoff, 1960)
2. Neocognitron (Fukushima, 1980)
3. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
4. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
5. Unsupervised convolutional network (Jarrett *et al.*, 2009)
6. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
7. Distributed autoencoder (Le *et al.*, 2012)
8. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
9. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
10. GoogLeNet (Szegedy *et al.*, 2014a)

Razões para Sucesso: Aumento do Tamanho de Modelos



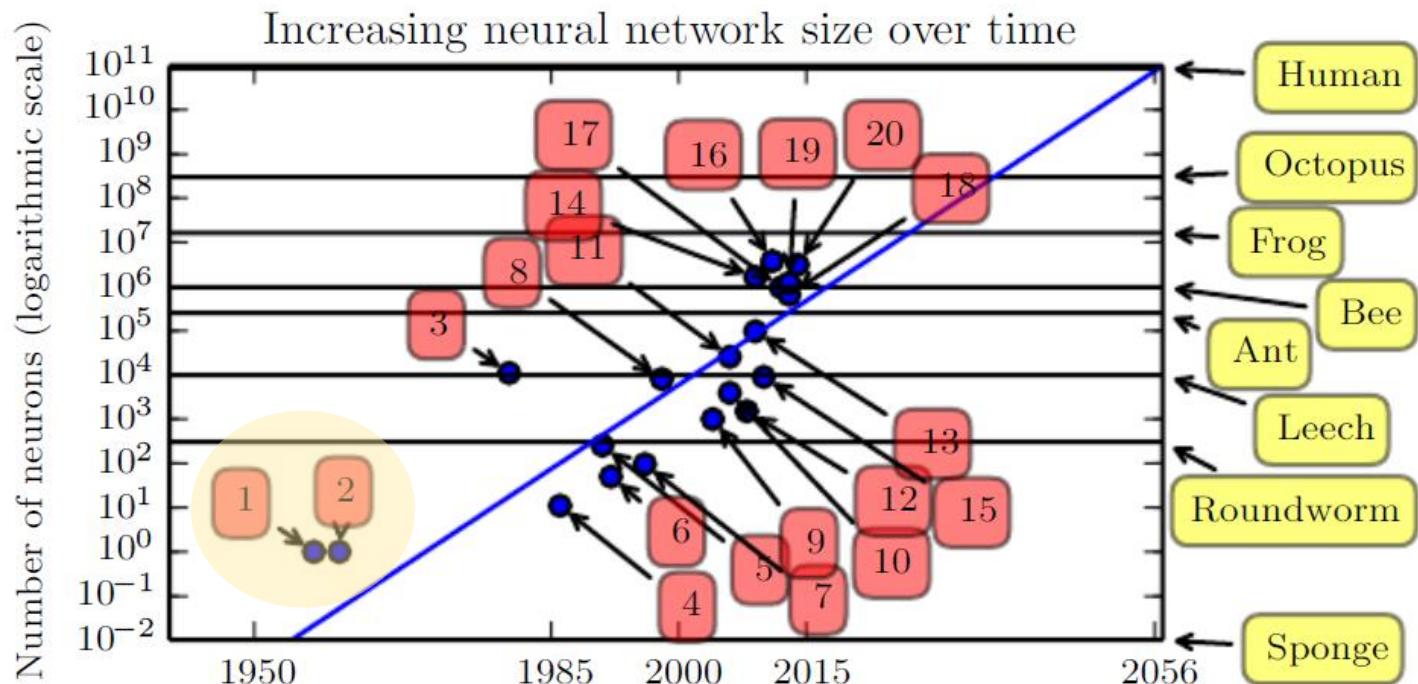
1. Adaptive linear element (Widrow and Hoff, 1960)
2. Neocognitron (Fukushima, 1980)
3. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
4. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
5. Unsupervised convolutional network (Jarrett *et al.*, 2009)
6. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
7. Distributed autoencoder (Le *et al.*, 2012)
8. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
9. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
10. GoogLeNet (Szegedy *et al.*, 2014a)

Razões para Sucesso: Aumento do Tamanho de Modelos



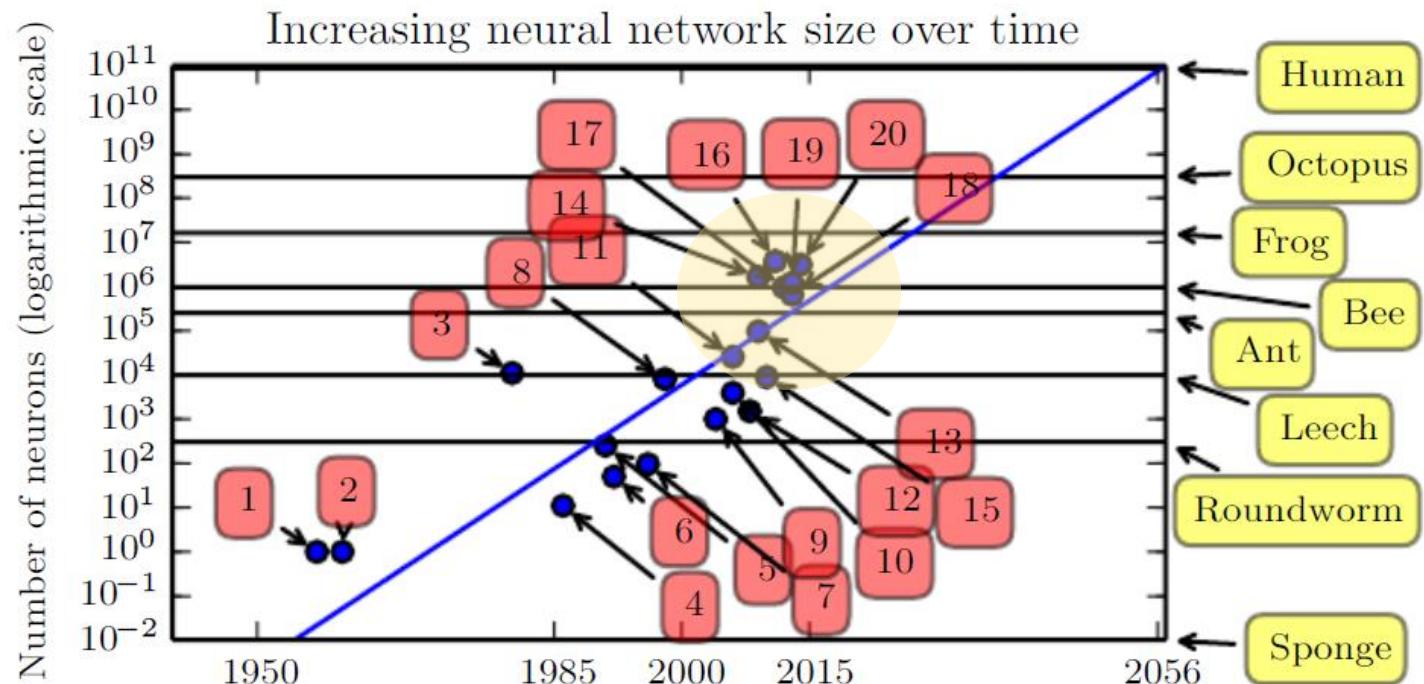
1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart *et al.*, 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallside, 1991)
6. Multilayer perceptron for speech recognition (Bengio *et al.*, 1991)
7. Mean field sigmoid belief network (Saul *et al.*, 1996)
8. LeNet-5 (LeCun *et al.*, 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton *et al.*, 2006)
11. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina *et al.*, 2009)
14. Unsupervised convolutional network (Jarrett *et al.*, 2009)
15. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le *et al.*, 2012)
18. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
19. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
20. GoogLeNet (Szegedy *et al.*, 2014a)

Razões para Sucesso: Aumento do Tamanho de Modelos



1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart et al., 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallside, 1991)
6. Multilayer perceptron for speech recognition (Bengio et al., 1991)
7. Mean field sigmoid belief network (Saul et al., 1996)
8. LeNet-5 (LeCun et al., 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton et al., 2006)
11. GPU-accelerated convolutional network (Chellapilla et al., 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina et al., 2009)
14. Unsupervised convolutional network (Jarrett et al., 2009)
15. GPU-accelerated multilayer perceptron (Ciresan et al., 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le et al., 2012)
18. Multi-GPU convolutional network (Krizhevsky et al., 2012)
19. COTS HPC unsupervised convolutional network (Coates et al., 2013)
20. GoogLeNet (Szegedy et al., 2014a)

Razões para Sucesso: Aumento do Tamanho de Modelos



1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart *et al.*, 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallside, 1991)
6. Multilayer perceptron for speech recognition (Bengio *et al.*, 1991)
7. Mean field sigmoid belief network (Saul *et al.*, 1996)
8. LeNet-5 (LeCun *et al.*, 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton *et al.*, 2006)
11. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina *et al.*, 2009)
14. Unsupervised convolutional network (Jarrett *et al.*, 2009)
15. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le *et al.*, 2012)
18. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
19. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
20. GoogLeNet (Szegedy *et al.*, 2014a)

Razões para Sucesso: Aumento da Acurácia

Aumento da acurácia \Rightarrow Aumento do impacto no mundo real

