

MBA  
USP  
ESALQ

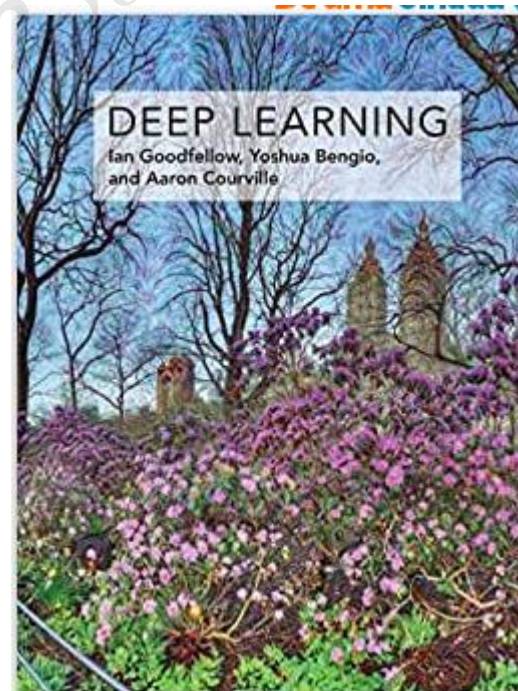
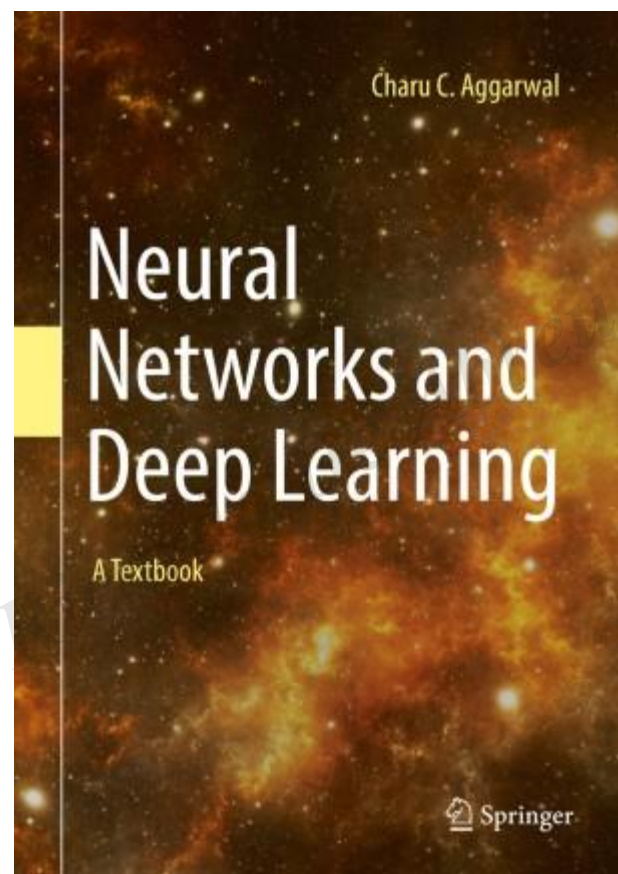
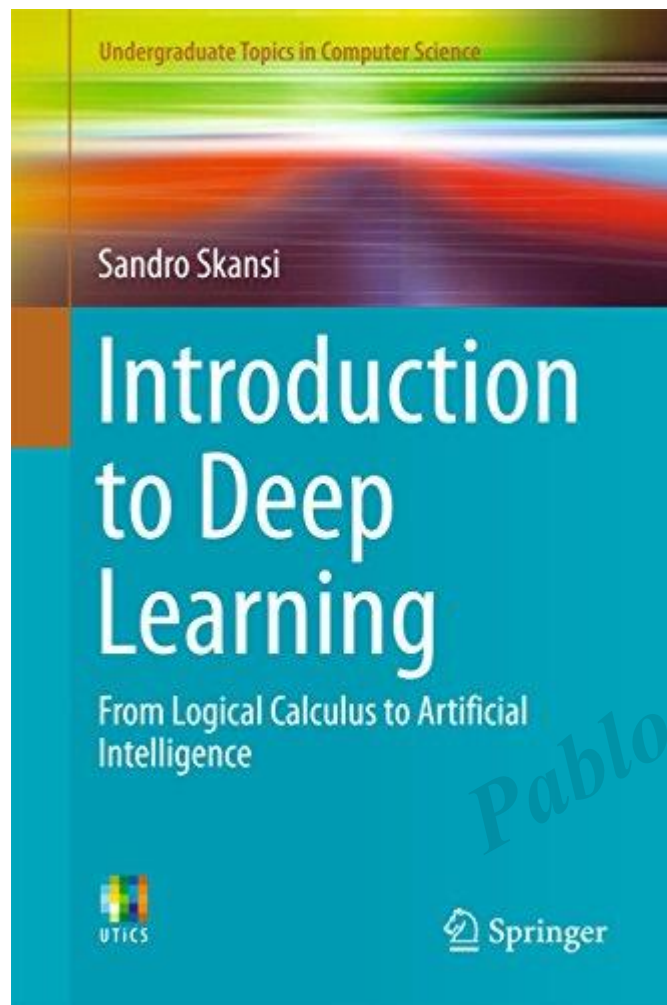
# DEEP LEARNING

Prof. Dr. Jeronimo Marcondes

**\*A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.**

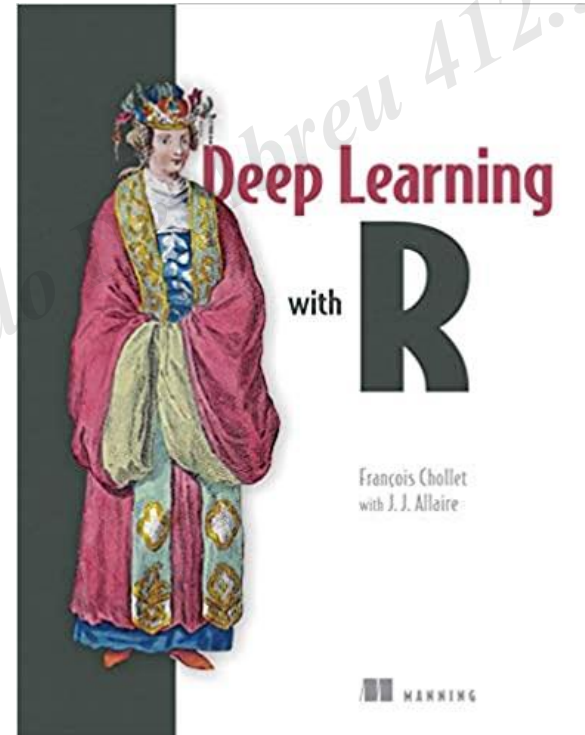
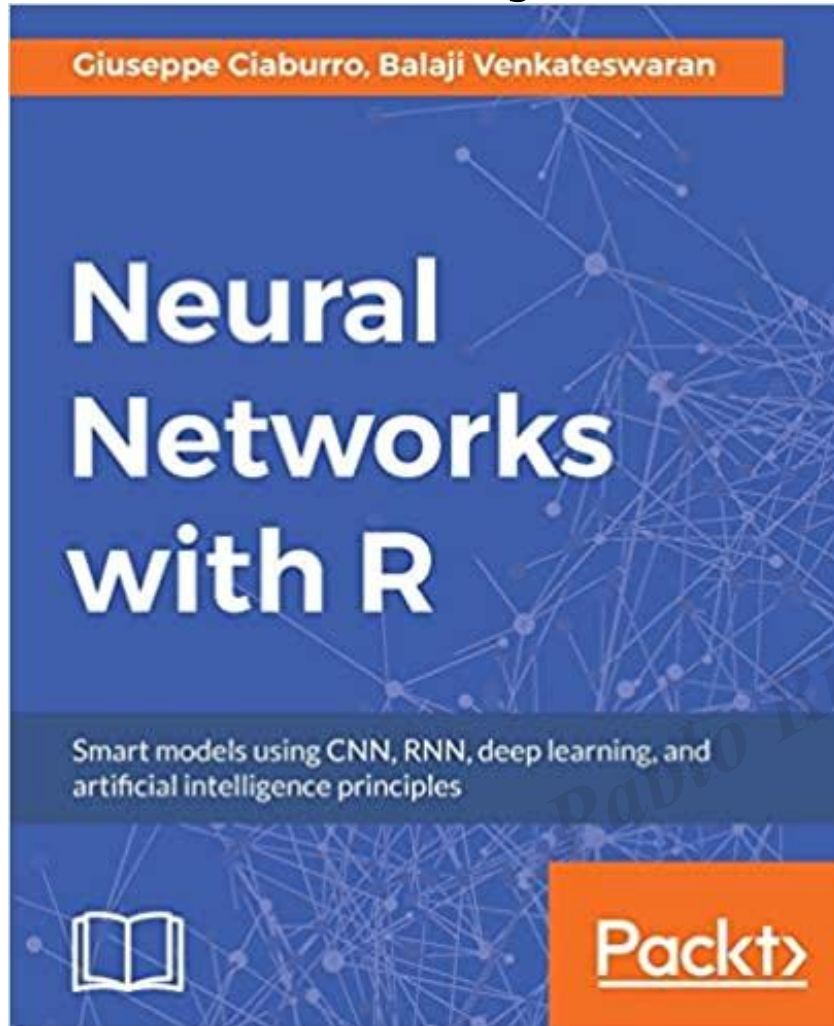
**Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98**

# Introdução





# Introdução



# Introdução

- Deep Learning não supervisionado
- O que é não supervisionado?
- Qual a diferença com relação aos modelos de FNN?

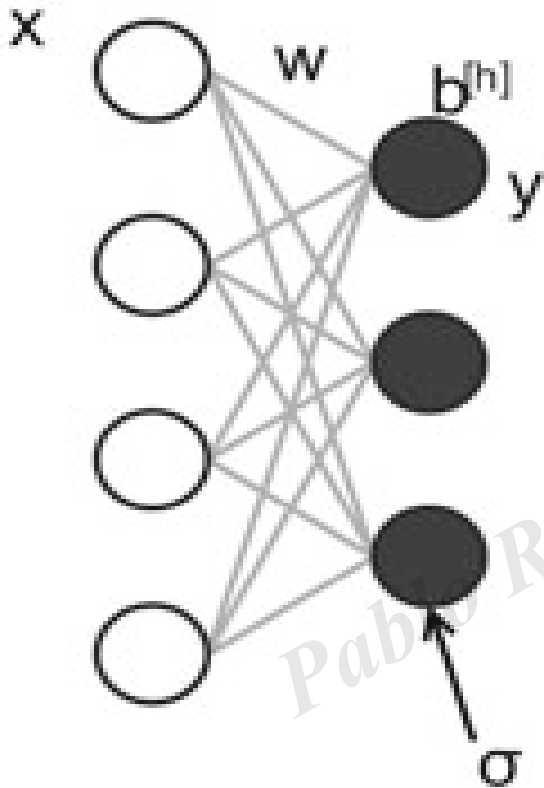
# Introdução

- Modelos Baseados em Energia
- Objetivo: Reduzir a energia
- Similar ao problema com nossa função custo

# Máquinas de Boltzmann

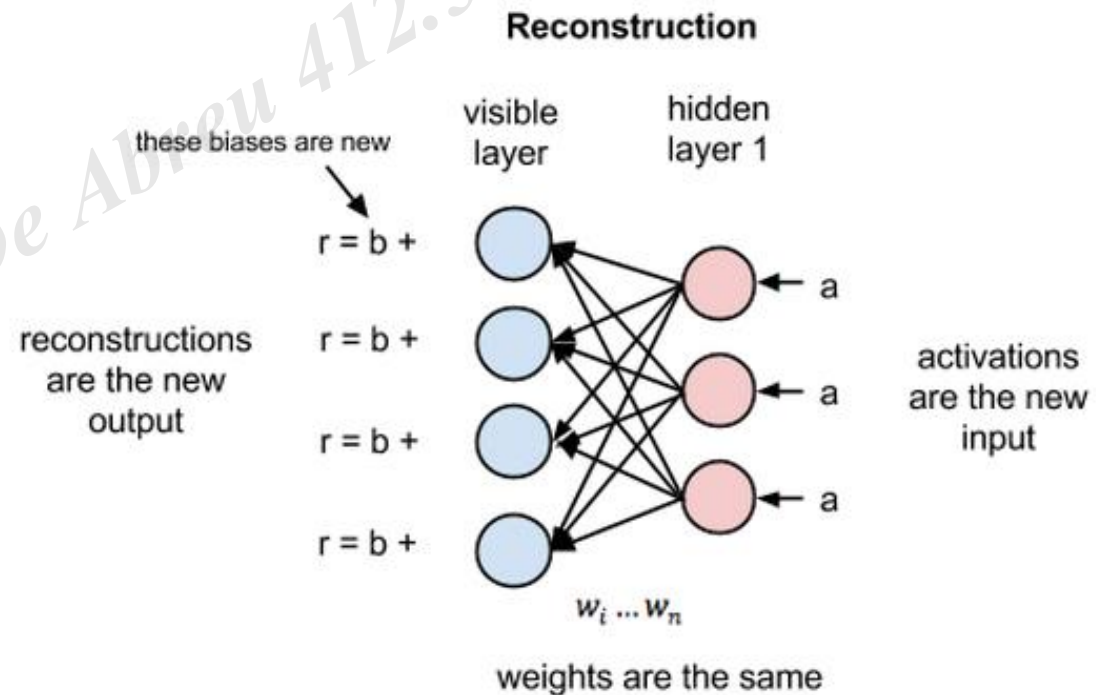
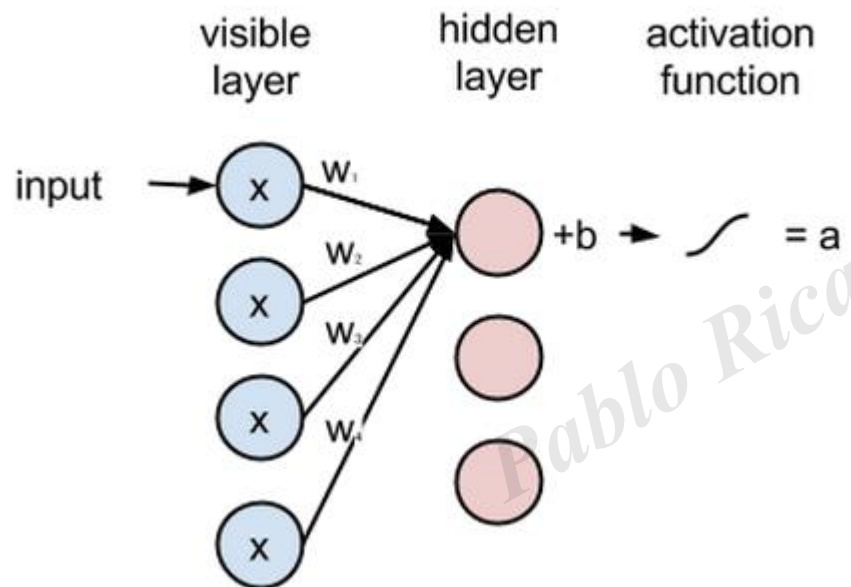
- A máquina de Boltzmann é uma FNN de uma camada
- Nosso objetivo: aprender a distribuição de probabilidades dos inputs
- Ajustar pesos para ser possível reconstruir os inputs (entradas)
- Máquina Restrita de Boltzmann

# Máquinas de Boltzmann





# Máquinas de Boltzmann



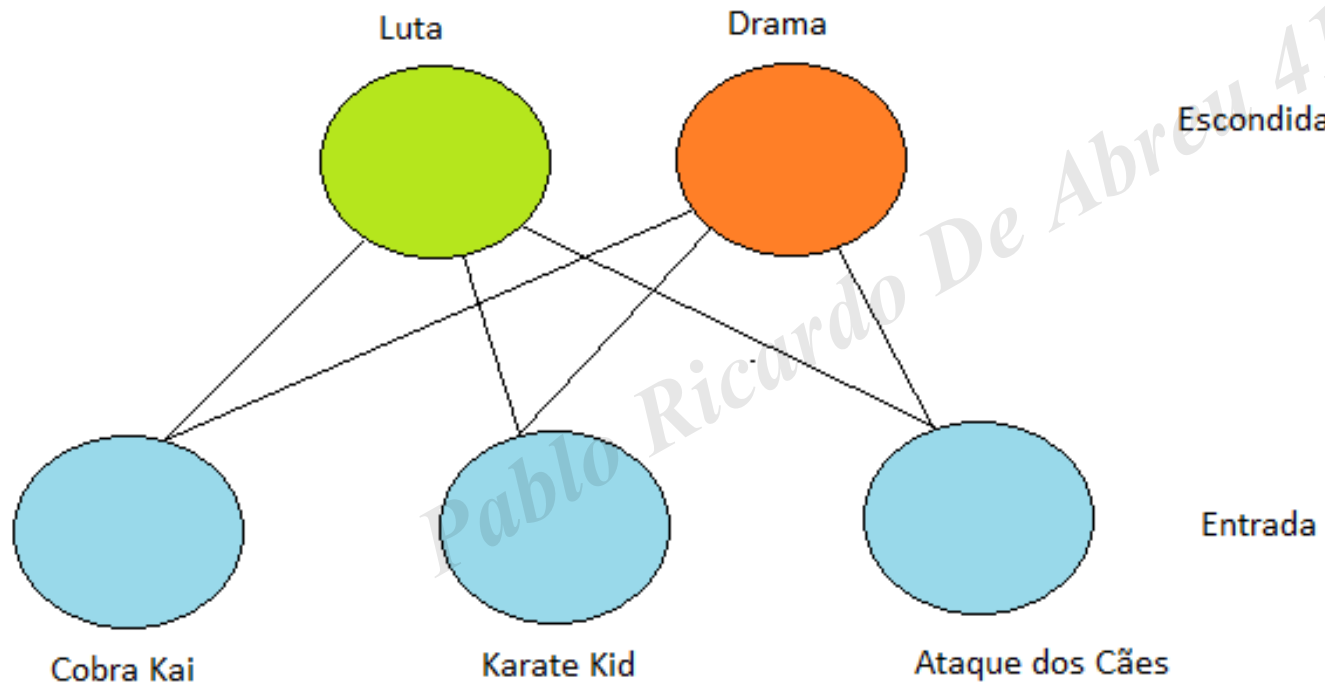
Fonte: <https://wiki.pathmind.com/restricted-boltzmann-machine>

# Máquinas de Boltzmann

Características:

- Não possui camada de saída
- Entradas são passadas para a camada escondida
- Por que é máquina “restrita”?

# Máquinas de Boltzmann



# Máquinas de Boltzmann

Funcionamento:

- Inicia camada interna randomicamente.
- Recebe insumo  $x$ .
- Calcula (probabilidade condicional que o neurônio seja ativado):

$$y = \sigma(xw + b_h)$$

# Máquinas de Boltzmann

Funcionamento:

- Y é devolvido para a camada interna para reconstrução.

- Calcula:

$$r = \sigma(yw + b_v)$$

- Intuição: variável latente

# Máquinas de Boltzmann

## Aprendizado pela Divergência Contrastiva

A divergência contrastiva é uma técnica de treinamento alternativa para aproximar a inclinação gráfica que representa a relação entre os pesos de uma rede e seu erro, chamada de gradiente. Como a maioria dos algoritmos de aprendizado probabilístico tenta otimizar o valor da probabilidade logarítmica, esse gradiente representa a direção desejada de mudança, de aprendizado, para os parâmetros da rede.

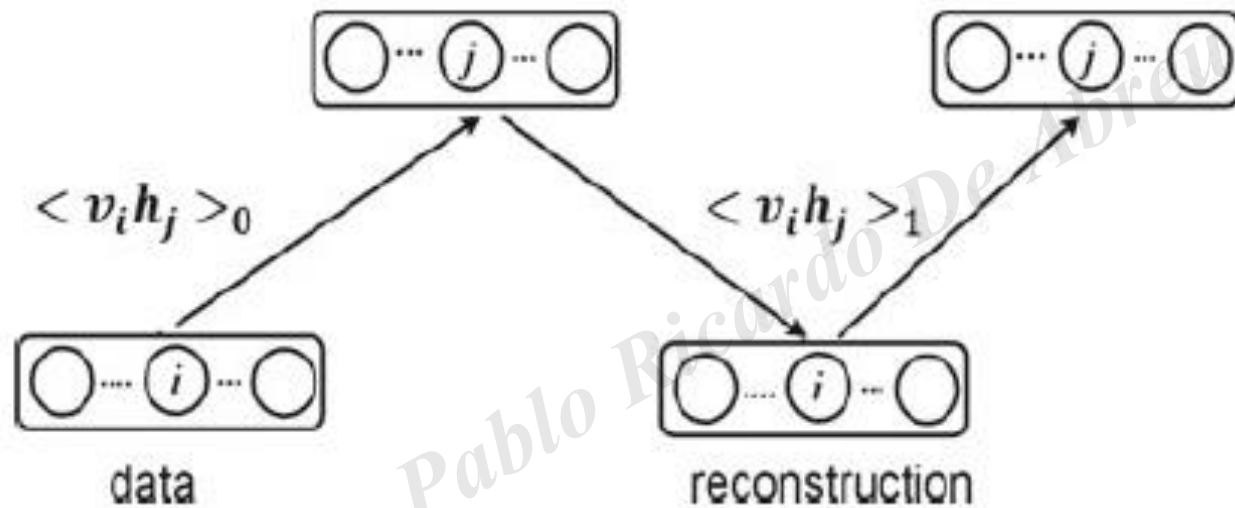


# Máquinas de Boltzmann

Funcionamento:

- A diferença entre entrada e  $r$  é verificada pela divergência de Kullback-Leibler.
- Fase negativa e positiva

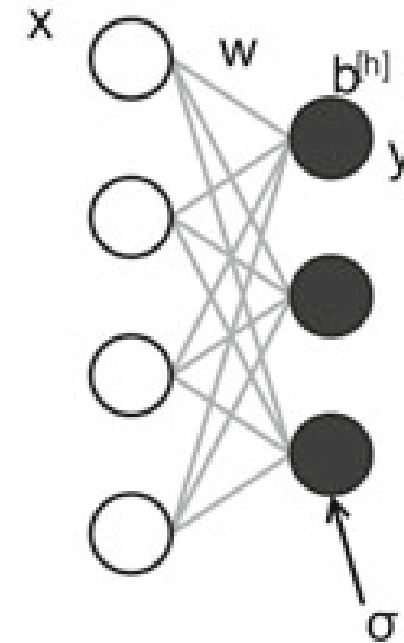
# Máquinas de Boltzmann



# Máquinas de Boltzmann

Aplicações:

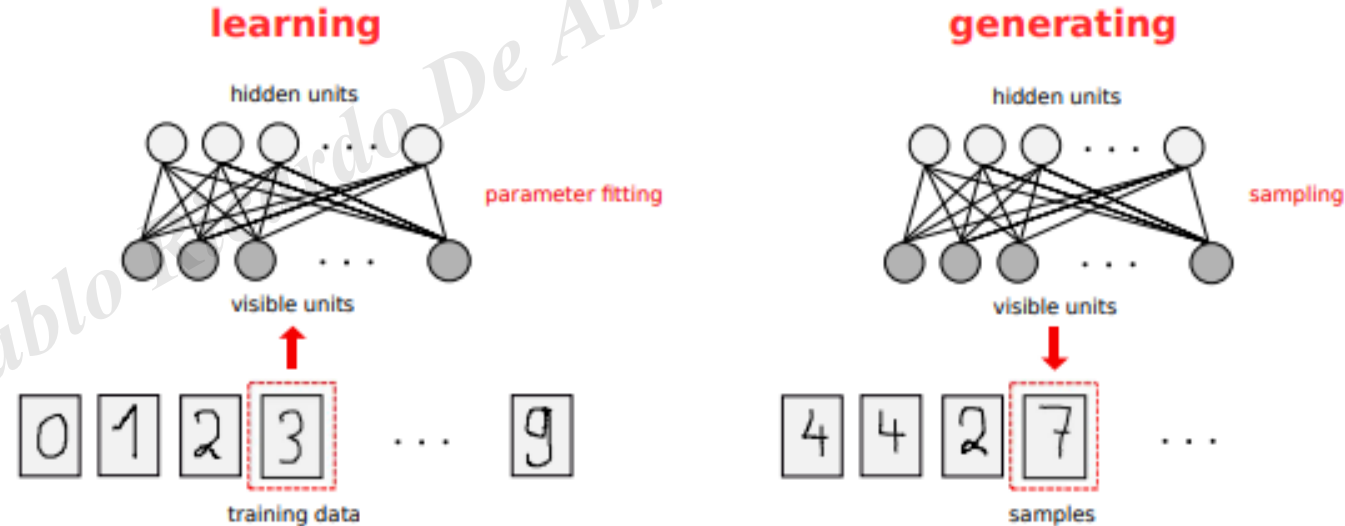
- Redução de Dimensionalidade
- Exemplo: problema com muitas variáveis.



# Máquinas de Boltzmann

Aplicações:

- Sistemas de recomendação
- Reconstrução de imagem



# Máquinas de Boltzmann

Exemplo: caso da NETFLIX

- Recomendação baseada em conteúdo – escolha semelhante ao que você já faz
- Recomendação baseada em filtragem colaborativa – perfis semelhantes

<https://tecnetit.com.br/como-a-netflix-usou-a-data-science-para-melhorar-seu-sistema-de-recomendacao/#:~:text=Para%20sua%20solu%C3%A7%C3%A3o%20eles%20usaram,dimensional%20ideal%20para%20seus%20usu%C3%A1rios.>

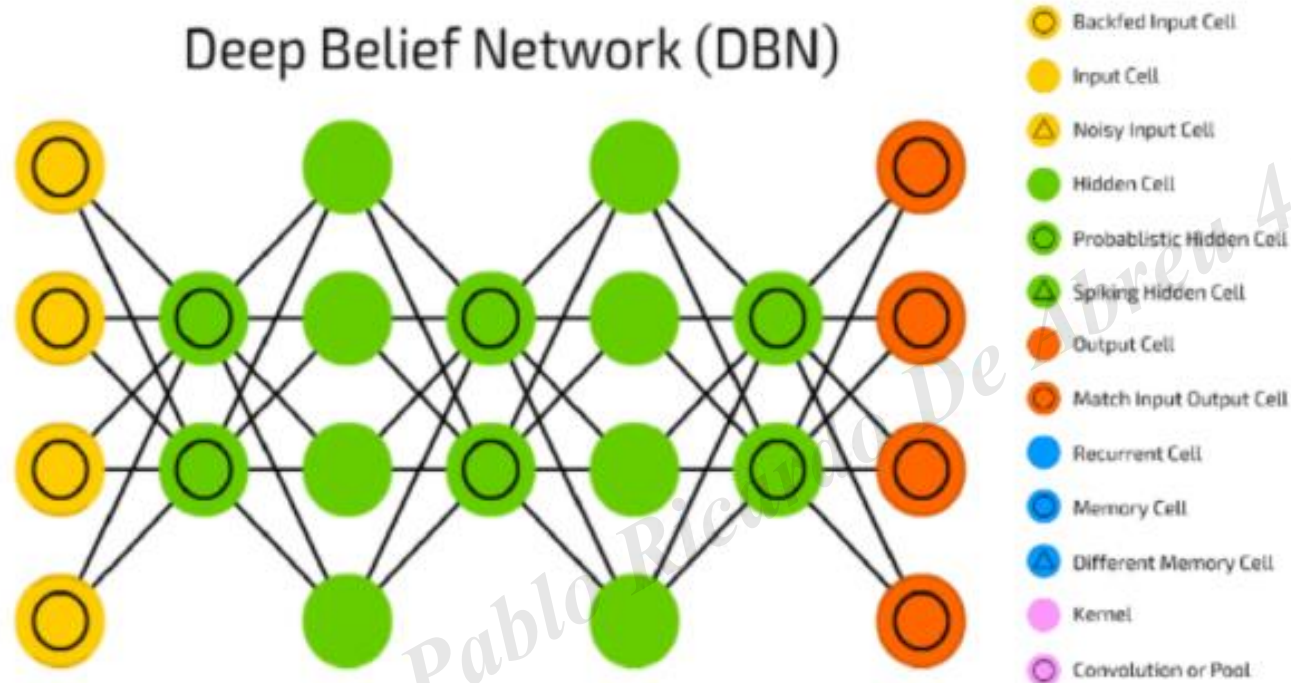
# DBN

## Deep Belief Networks

- Caso mais generalizado da Máquina restrita de Boltzmann.
- Empilhar máquinas restritas de Boltzmann



# DBN

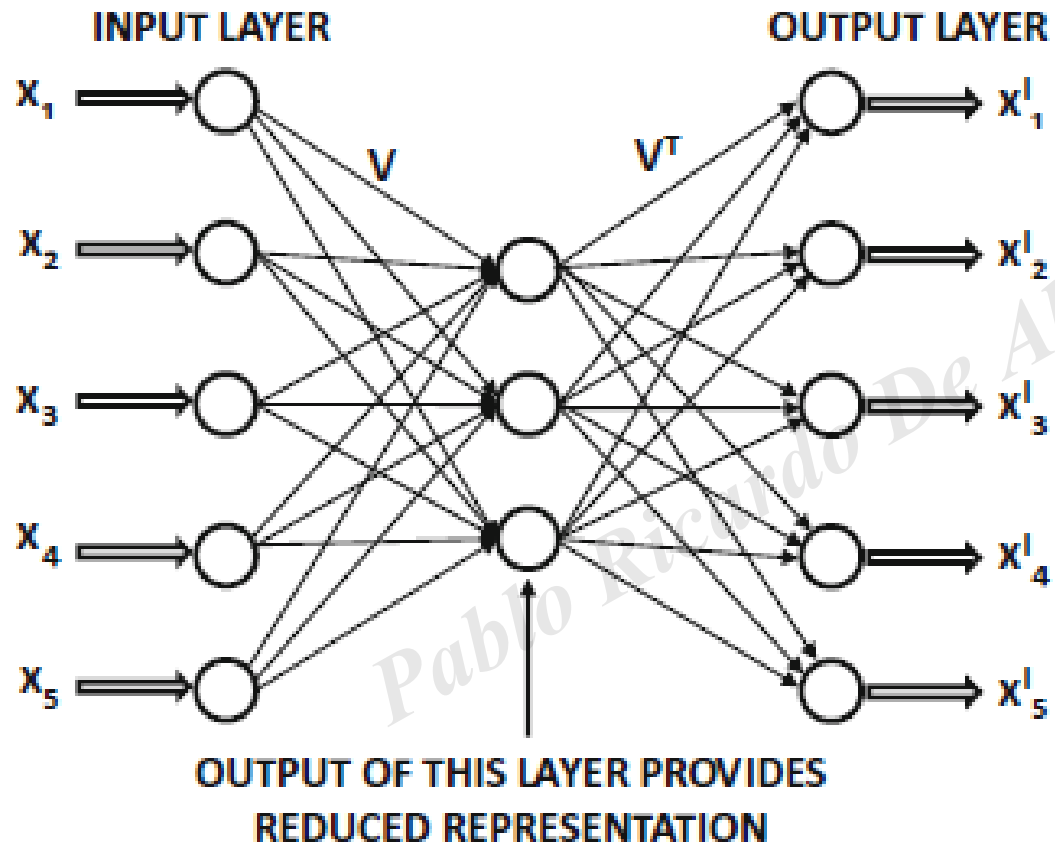


Fonte: <https://www.deeplearningbook.com.br/>

# Autoencoder

- Objetivos muito semelhantes à Máquina de Boltzmann
- Reconstrução de input
- Redução de dimensionalidade

# Autoencoder



# Autoencoder

- Autoencoder é como um funil
- Passa a informação e a restringe em um número menor de camadas
- Após isso tentamos reconstruir a informação

# Autoencoder

- Etapa de “encode” – codifica a imagem em uma dimensionalidade menor:

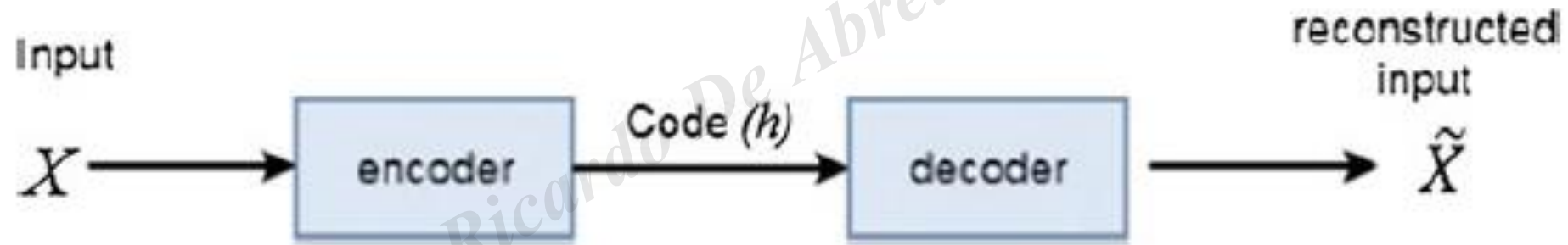
$$h = f(x)$$

- Etapa de “decode” – decodifica a imagem de forma a reconstruir o input.

$$r = f(h)$$

- Diferença com relação à máquina restrita de boltzmann

# Autoencoder



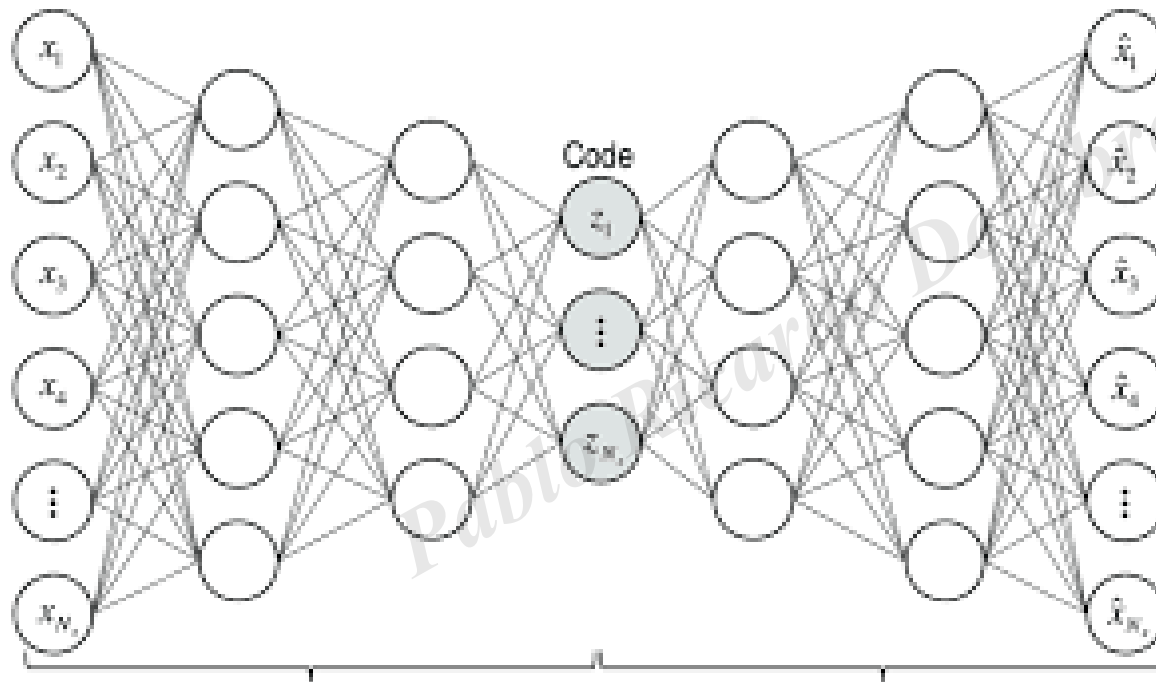


# Autoencoder

<https://douglasduhaime.com/posts/visualizing-latent-spaces.html>

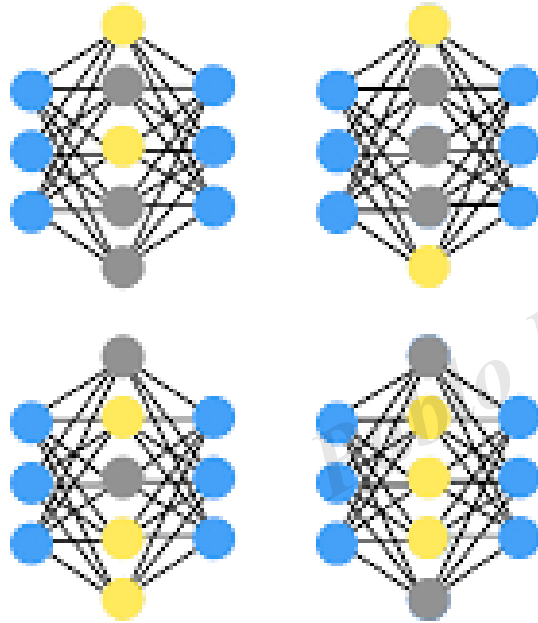
# Tipos de Autoencoder

- Deep Autoencoder



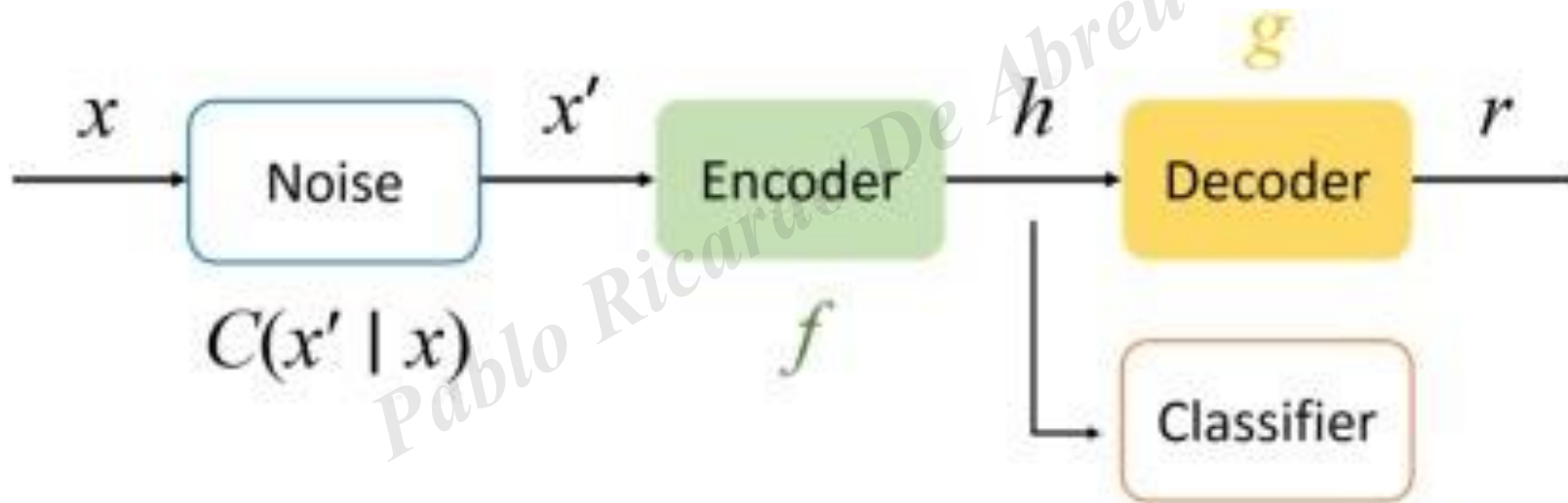
# Tipos de Autoencoder

- Autoencoder Esparso
- Aplica-se termo de regularização



# Tipos de Autoencoder

- Autoencoder Denoising



# Tipos de Autoencoder

- Autoencoder Contrativo
- Diferença com relação ao denoising — inclui penalidade na função custo
- Denoising inclui nos dados de treinamento
- Obtem resultados mais interessantes que o denoising

# GAN

- Redes Adversárias Generativas
- “a ideia mais interessante nos últimos 10 anos em Machine Learning”
- Duas redes competindo
- Modelo Generativo



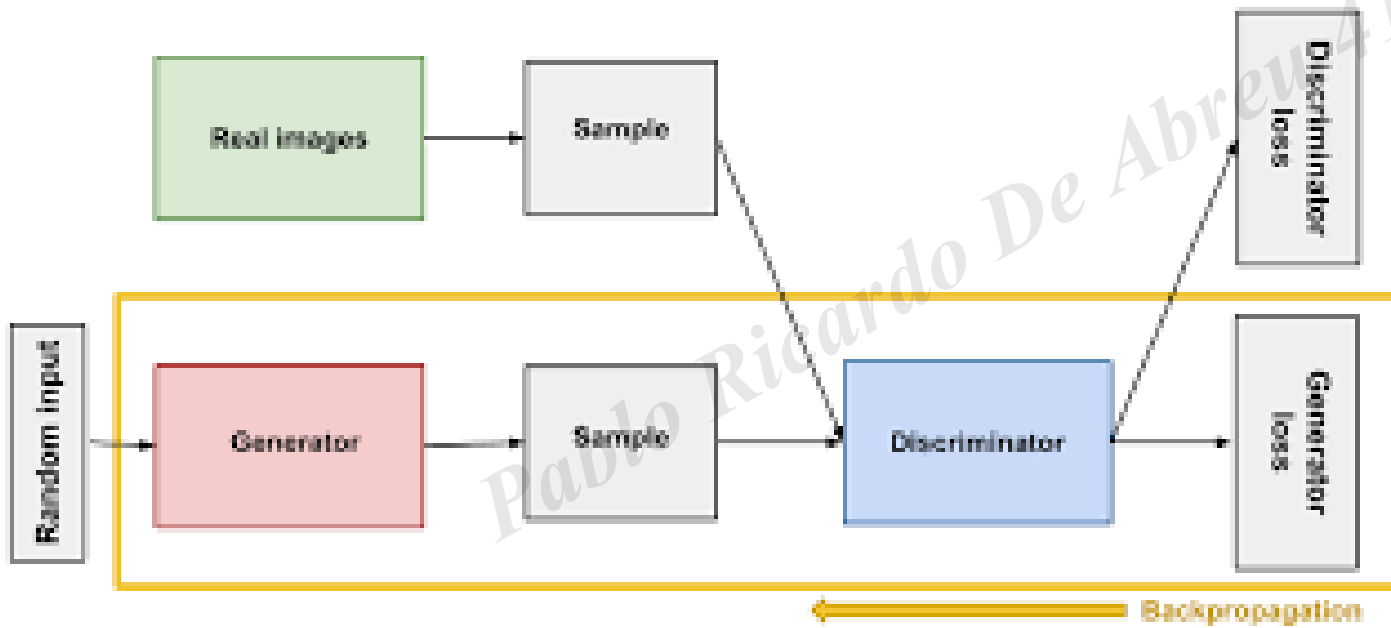
# GAN



# GAN

- Como obter  $y$  dado  $x$ ?
- Modelos generativos modelam a distribuição de classes individuais.
- Modelos discriminativos aprendem a fronteira entre classes.

# GAN



# GAN

- Exemplo MNIST
- Gerador irá gerar imagens “fake”
- Passaremos imagens reais junto com as geradas
- O discriminador deve reconhecer as imagens e fazer a diferença

# Parte prática



# GAN

- Realizamos o backpropagation
- O gerador irá gerar imagens cada vez melhores
- O discriminador deverá ficar cada vez melhor em discriminar
- O discriminador obtém imagens reais e falsas e retorna probabilidades, um número entre 0 e 1, com 1 representando uma previsão de imagem autêntica e 0 representando previsão de imagens falsas (geradas pela rede generativa).

# Discussão

*Pablo Ricardo De Abreu 412.566.518-42*

**OBRIGADO!**

<https://www.linkedin.com/in/jeronymo-marcondes-585a26186>