

# Autoencoders

Eduardo de Medeiros da Silveira

Universidade Federal de Santa Maria

# Representação Eficiente de Dados

Em 1970, William Chase and Herbert Simon fizeram um experimento com jogadores profissionais de xadrez, para estudar a relação entre memória, percepção e reconhecimento de padrões.

- ▶ Capazes de memorizar o tabuleiro em poucos segundos.
- ▶ Somente quando as peças estavam em posições naturais.
- ▶ O reconhecimento de padrões ajuda na memorização.

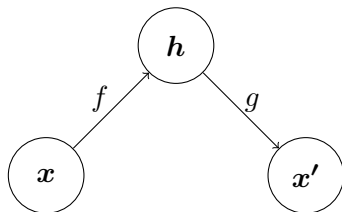
# Representação Eficiente de Dados



Figura: Etapas do experimento da memória no xadrez.

# Autoencoder

Um *autoencoder* é uma rede neural que **tenta** aprender a função identidade.



**Figura:** Esquema geral de um *autoencoder*, que mapeia uma entrada  $x$  para uma saída  $x'$ , através de uma representação interna  $h$ . O *autoencoder* é composto por um codificador  $f$  e um decodificador  $g$ .

# Autoencoder

Algumas características:

- ▶ Aprendizado não-supervisionado ou auto-supervisionado.
- ▶ A saída não importa.
- ▶ A representação latente  $\mathbf{h}$  importa.
- ▶ Restrições.

# Autoencoder

Algumas características:

- ▶ Mesmo número de neurônios na entrada e na saída.
- ▶ Geralmente é simétrico.

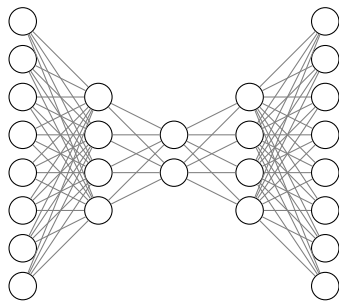


Figura: Exemplo de um autoencoder.

# Undercomplete Autoencoder

- ▶ A dimensão de  $\mathbf{h}$  é menor do que a dimensão de  $\mathbf{x}$ .
- ▶ Minimiza-se  $L$ , que calcula a dissimilaridade de  $\mathbf{x}$  e  $\mathbf{x}'$ .
- ▶ Não há garantia de que  $\mathbf{h}$  capture informações importantes.

$$L(\mathbf{x}, \mathbf{x}') = L(\mathbf{x}, g(f(\mathbf{x})))$$