

Autoencoders: da motivação às variantes modernas

CPE 727 - Aprendizado de Profundo

Felipe Fink Grael, Rafael Tadeu Cardoso dos Santos, Thalles Nonato Leal Santos e Jefferson Osowsky

18 de novembro de 2025

Table of Contents

1 Motivação

► Motivação

► Formulação Matemática

► Referências Bibliográficas

Introdução

1 Motivação

- **Definição:** Algoritmos cujo propósito principal é copiar sua entrada na saída [1]. São tipicamente construídos como redes neurais artificiais treinadas de forma não supervisionada.
- **Arquitetura Básica:**
 - Encoder: transforma entrada em representação latente
 - Representação: espaço latente de menor, maior ou igual dimensão
 - Decoder: reconstrói a entrada original
- **Objetivo:** Aprender a função identidade $f(x) \approx x$ através de um espaço latente comprimido

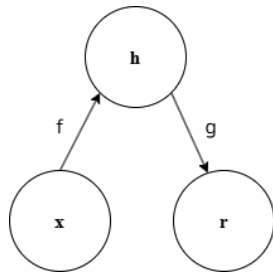


Figura: Estrutura básica de Autoencoders.

Componentes Fundamentais

1 Motivação

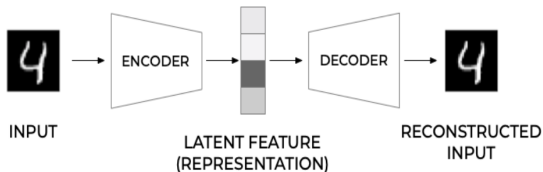


Figura: Estrutura de um autoencoder com representação Latente¹.

- **Encoder:** $f_{\theta} : \mathcal{X} \rightarrow \mathcal{H}$ onde $h = f_{\theta}(x)$
- **Decoder:** $g_{\phi} : \mathcal{H} \rightarrow \mathcal{X}$ onde $x' = g_{\phi}(h)$
- **Reconstrução:** $x' = g_{\phi}(f_{\theta}(x))$
- **Espaço latente \mathcal{H} :** Representação comprimida dos dados ($\dim(\mathcal{H}) < \dim(\mathcal{X})$)

¹Figura de Umberto Michelucci [2]

Por que usar Autoencoders?

1 Motivação

Aprendizado de Representações:

- Extrair características relevantes automaticamente dos dados
- Redução de dimensionalidade não-linear (superior ao PCA para dados complexos)
- Aprendizado não supervisionado - não requer labels

Vantagens sobre métodos tradicionais:

- PCA: apenas transformações lineares
- Autoencoders: capturam relações não-lineares complexas
- Profundidade permite representações hierárquicas [3]

Table of Contents

2 Formulação Matemática

► Motivação

► **Formulação Matemática**

► Referências Bibliográficas

Definição Formal

2 Formulação Matemática

Seja μ_{ref} uma distribuição de probabilidade de referência em \mathcal{X} e $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ uma função de distância.

Função de custo do autoencoder:

$$L(\theta, \phi) = \mathbb{E}_{x \sim \mu_{ref}} [d(x, g_{\phi}(f_{\theta}(x)))]$$

Objetivo de treinamento:

$$(\theta^*, \phi^*) = \arg \min_{\theta, \phi} L(\theta, \phi)$$

Exemplo: Autoencoder Linear de Uma Camada

2 Formulação Matemática

Encoder:

$$h = f_{W,b}(x) = \sigma(Wx + b)$$

Decoder:

$$x' = g_{W',b'}(h) = \sigma(W'h + b')$$

Função Custo:

$$L(W, b, W', b') = \frac{1}{N} \sum_{i=1}^N \|x_i - g_{W',b'}(f_{W,b}(x_i))\|_2^2$$

Parâmetros a otimizar: $\theta = W, b, \phi = W', b'$

Exemplo: Undercomplete Autoencoder Linear (Caso Especial)

2 Formulação Matemática

Autoencoder linear: sem função de ativação $\sigma(z) = z$

Teorema: O autoencoder linear ótimo projeta os dados no subespaço gerado pelos primeiros k autovetores da matriz de covariância Σ_{XX} [4].

Erro mínimo:

$$\Sigma(A, B) = \text{Tr}(\Sigma) - \sum_{i=1}^k \lambda_i = \sum_{i=k+1}^n \lambda_i$$

Onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ são os autovalores de Σ_{XX} .

Conexão com PCA: Autoencoders lineares aprendem o mesmo subespaço que a Análise de Componentes Principais.

Table of Contents

3 Referências Bibliográficas

► Motivação

► Formulação Matemática

► Referências Bibliográficas

Referências Bibliográficas

3 Referências Bibliográficas

- [1] Rumelhart, E. David, M. James, and L. James, *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations*.
 O1 1986.
- [2] U. Michelucci, “An introduction to autoencoders,” *CoRR*, vol. abs/2201.03898, 2022.
- [3] M. Tschannen, O. Bachem, and M. Lucic, “Recent advances in autoencoder-based representation learning,” *CoRR*, vol. abs/1812.05069, 2018.
- [4] E. Oja, “Simplified neuron model as a principal component analyzer,” *Journal of Mathematical Biology*, vol. 15, pp. 267–273, 1982.

Autoencoders: da motivação às variantes modernas

Obrigado pela Atenção!

Alguma Pergunta?

Natanael Moura Junior

natmourajr@poli.ufrj.br, natmourajr@lps.ufrj.br