Aula 8: Autoencoders Variacionais (VAEs) I

prof. Ranniery Maia

DIM0095 - Tópicos Especiais em Computação VI
(IA Geradora)

28 de Abril de 2025

Conteúdo

Introdução aos VAEs

- Autoencoders Variacionais (VAEs) são uma extensão dos autoencoders tradicionais
- Desenvolvidos por Kingma e Welling (2013)
- Combinam aprendizado não-supervisionado com modelagem probabilística
- Diferente dos autoencoders regulares, VAEs são modelos generativos
- Permitem gerar novas amostras, não apenas reconstruir

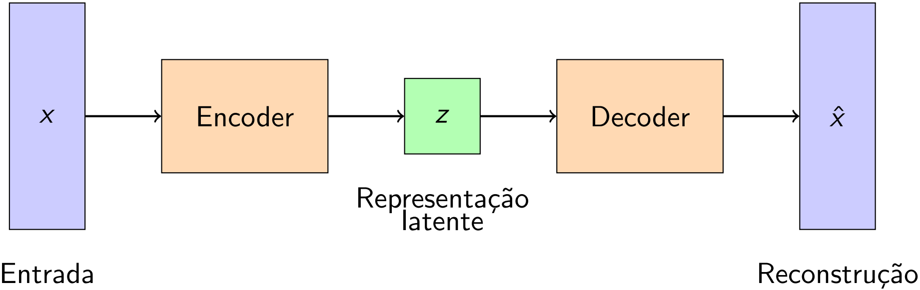
Autoencoders Tradicionais vs. VAEs

Autoencoder Tradicional:

- Codifica entradas em vetores de características fixos
- Decodifica para reconstruir a entrada original
- Espaço latente descontínuo e não estruturado

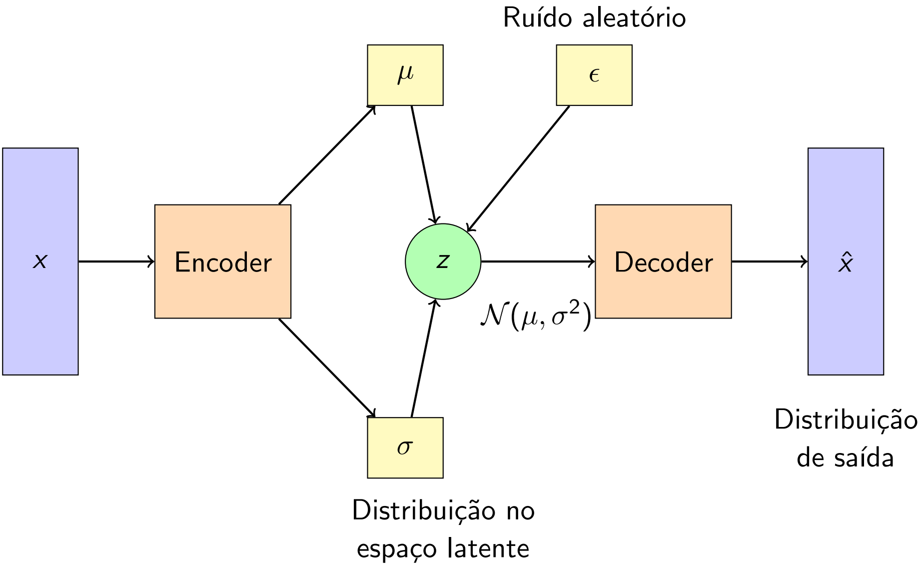
Autoencoder Variacional:

- Codifica entradas em distribuições probabilísticas
- Decodifica amostras dessas distribuições
- Espaço latente contínuo e bem estruturado
- Permite geração de novos dados



No autoencoder tradicional, a entrada é codificada em um único ponto no espaço latente.

Autoencoder Variacional (VAE)



Autoencoders Tradicionais vs. VAEs

Autoencoder Tradicional:

- Codifica entradas em vetores de características fixos
- Decodifica para reconstruir a entrada original
- Espaço latente descontínuo e não estruturado

Autoencoder Variacional:

- Codifica entradas em distribuições probabilísticas
- Decodifica amostras dessas distribuições
- Espaço latente contínuo e bem estruturado
- Permite geração de novos dados

Teoria Probabilística dos VAEs

- VAEs são baseados em inferência variacional
- Objetivo: modelar a distribuição dos dados $p(x)$
- Introduz variáveis latentes z para modelar $p(x) = \int p(x|z)p(z)dz$
- Problema: esta integral é geralmente intratável
- Solução: aproximar a distribuição posterior $p(z|x)$ com uma distribuição $q_\phi(z|x)$ mais simples

Componentes Principais do VAE

- **Encoder:** Rede neural que aproxima $q_\phi(z|x)$ (distribuição posterior)
- **Distribuição Anterior (a priori):** Geralmente escolhida como $p(z) = \mathcal{N}(0, I)$
- **Decoder:** Rede neural que modela $p_\theta(x|z)$ (verossimilhança)
- O modelo completo é treinado end-to-end

Modelagem Matemática

O VAE resolve o problema:

$$\log p(x) = \log \int p(x|z)p(z)dz \tag{1}$$

$$\geq \mathbb{E}_{z\sim q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(z)) \tag{2}$$

Onde:

- $\log p(x)$ é o log da verossimilhança dos dados
- D_{KL} é a divergência Kullback-Leibler
- Esta desigualdade é conhecida como ELBO (Evidence Lower Bound)

Espaço Latente em VAEs

- No autoencoder tradicional: pontos discretos no espaço latente
- No VAE: regiões contínuas (distribuições) no espaço latente
- Cada entrada x é mapeada para uma distribuição $q_\phi(z|x)$
- Tipicamente modelada como uma distribuição gaussiana multivariada
- O encoder produz parâmetros desta distribuição (μ e σ)

Modelagem do Espaço Latente

- O encoder produz $\mu(x)$ e $\sigma(x)$ para cada entrada x
- Definimos $q_\phi(z|x) = \mathcal{N}(z; \mu(x), \sigma^2(x)I)$
- Durante o treinamento, amostramos $z \sim q_\phi(z|x)$
- Durante a geração, amostramos $z \sim p(z) = \mathcal{N}(0, I)$
- O espaço latente organiza dados de forma significativa

Propriedades do Espaço Latente

- **Continuidade:** Pontos próximos no espaço latente geram dados similares
- **Completude:** Qualquer ponto no espaço latente gera dados realistas
- **Disentanglement:** Diferentes dimensões capturam fatores independentes de variação
- Permite interpolação e exploração de dados no espaço latente
- Facilita tarefas como geração, edição e transformação

Reparametrização

- Problema: Como propagar gradientes através de um processo de amostragem?
- Solução: *truque de reparametrização*
- Em vez de amostrar diretamente $z \sim \mathcal{N}(\mu, \sigma^2)$, fazemos:
- Amostramos $\epsilon \sim \mathcal{N}(0, I)$
- Computamos $z = \mu + \sigma \odot \epsilon$
- Isto permite o backpropagation através da operação de amostragem

Função de Perda do VAE

A função de perda do VAE tem duas componentes:

$$\mathcal{L}(\theta, \phi; x) = \mathcal{L}_{\text{reconstrução}} + \mathcal{L}_{\text{regularização}} \tag{3}$$

$$= -\mathbb{E}_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)] + D_{KL}(q_\phi(z|x) || p(z)) \tag{4}$$

- $\mathcal{L}_{\text{reconstrução}}$: Assegura que os dados decodificados sejam similares aos originais
- $\mathcal{L}_{\text{regularização}}$: Força $q_\phi(z|x)$ a ser próxima de $p(z)$

Termo de Reconstrução

- $\mathcal{L}_{\text{reconstrução}} = -\mathbb{E}_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)]$
- Mede quão bem o modelo pode reconstruir a entrada original
- Para dados de imagem binários: perda de entropia cruzada binária
- Para dados contínuos: erro quadrático médio ou distribuição gaussiana
- Na prática, estimamos usando L amostras: $\frac{1}{L}\sum_{l=1}^{L} \log p_\theta(x|z^{(l)})$

Termo de Regularização

- $\mathcal{L}_{\text{regularização}} = D_{KL}(q_\phi(z|x)||p(z))$
- Força a distribuição posterior a ser próxima da distribuição a priori
- Para $q_\phi(z|x) = \mathcal{N}(\mu, \sigma^2)$ e $p(z) = \mathcal{N}(0, I)$:
- $D_{KL} = \frac{1}{2}\sum_{j=1}^{J}(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2)$
- Encoraja distribuições latentes suaves e bem comportadas

Balanceamento dos Termos da Perda

- O balanceamento entre reconstrução e regularização é crucial
- Muito peso na reconstrução: boa reconstrução, mas espaço latente desorganizado
- Muito peso na regularização: espaço latente bem organizado, mas reconstruções pobres
- VAE-β: introduz um hiperparâmetro β para controlar este equilíbrio
- $\mathcal{L}(\theta, \phi; x) = -\mathbb{E}[\log p_\theta(x|z)] + \beta \cdot D_{KL}(q_\phi(z|x)||p(z))$

Desafios e Soluções

- **Posterior collapse:** Quando o decodificador ignora o espaço latente
- **Lacuna em reconstruções:** VAEs podem gerar reconstruções borradas
- **Soluções:**
  - Arquiteturas mais complexas (ConvVAE, ResVAE)
  - Distribuições a priori mais expressivos (VampPrior, FlowPrior)
  - Funções de perda modificadas (VAE-β, InfoVAE)

Resumo

- VAEs combinam redes neurais com modelagem probabilística
- O espaço latente contínuo permite geração e interpolação
- A função de perda equilibra reconstrução e regularização
- VAEs são fundamentais para muitas aplicações modernas:
  - Geração de imagens, texto e música
  - Compressão de dados
  - Detecção de anomalias
  - Aprendizado semi-supervisionado

Próxima Aula: VAEs II

Na próxima aula, abordaremos:

- Implementação prática de VAEs
- Variantes avançadas de VAEs
- Aplicações e exemplos
- Comparação com outros modelos generativos

Referências

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... & Lerchner, A. (2017). beta-VAE: Learning basic visual concepts with a constrained variational framework.