

# Deep Learning - Relatório 4:

## *Autoencoders e GANs*

Rian Freitas e Hanna Rodrigues

Novembro 2023

## 1 Introdução

Neste trabalho, investigamos e comparamos o desempenho dos modelos Autoencoders e GANs na tarefa de síntese de dados usando o conjunto de dados Fashion MNIST. O Fashion MNIST é um conjunto de dados semelhante ao MNIST, mas em vez de dígitos manuscritos, ele consiste em imagens em escala de cinza de itens de moda. Este conjunto de dados inclui 70.000 imagens de vestuário, calçados e acessórios, cada uma com uma resolução de 28x28 pixels. Para nossos experimentos, utilizaremos 60.000 imagens para o treinamento e 10.000 para o conjunto de teste.

## 2 Autoencoder (AE)

O primeiro modelo utilizado foi um autoencoder simples, AE, com camadas densas e função de ativação sigmoid para o decoder e relu para a camada latente. Foram utilizados os seguintes hiperparâmetros:

- Dimensão do espaço latente : 2 e 64
- Epochs : 10
- Otimizador : Adam
- Função de Perda : MSE
- batch size : 32

Para o encoder de dimensão latente igual a 2 o treinamento para todo os dados do fashion MNIST levou 82.94 segundos. E as perdas tanto para o treinamento quanto para a validação foram ao longo das épocas em 0.044, fazendo-nos inferir que o modelo não aprendeu durante o treinamento.

Ao treinar novamente com 64 dimensões latentes o treinamento levou 83.41 segundos e ao longo das epochs a loss diminuiu, ou seja, o modelo aprendeu durante o treinamento. Isso nos faz inferir que o encoder requer uma espaço de

dimensões latentes maior para capturar as características intrínsecas do nosso dado de vestuário.

A seguir plotamos imagens do conjunto de teste e suas respectivas versões reconstruídas pelos autoencoder de dimensão latente igual a 2 e a 64. Novamente conseguimos ver que o aumento das dimensões espaço latente proporcionou uma reconstrução mais detalhada as imagens tanto que para os dados de teste quanto de treino.

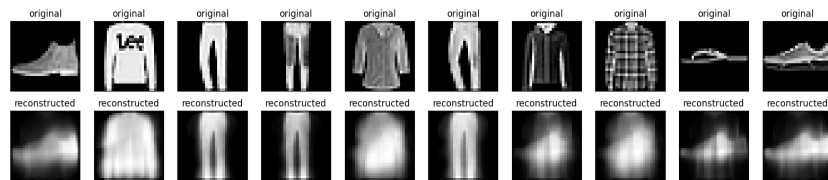


Figure 1: autoencoder com 2 dimensões latentes nos dados de teste

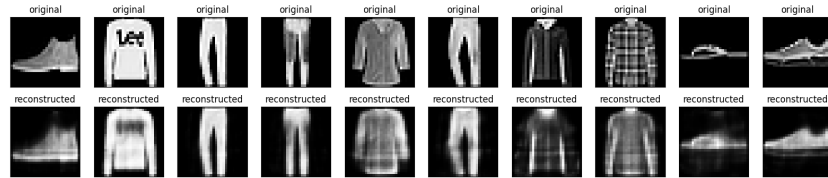


Figure 2: autoencoder com 64 dimensões latentes nos dados de teste

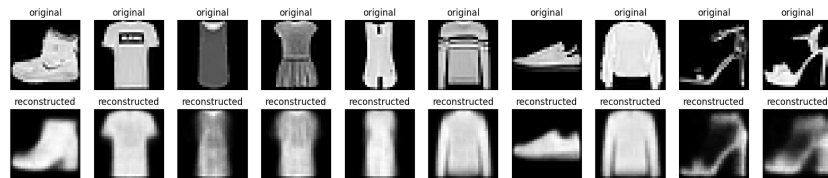


Figure 3: autoencoder com 2 dimensões latentes nos dados de treino

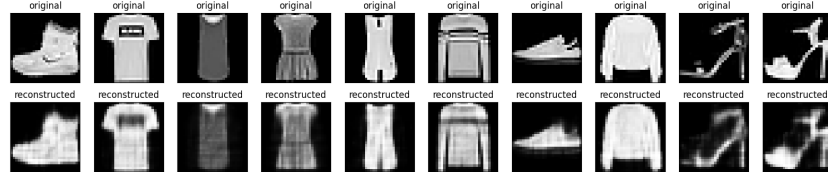


Figure 4: autoencoder com 64 dimensões latentes nos dados de treino

Para o autoencoder de espaço latente bidimensional conseguimos visualizar a seguinte matriz de reconstrução, podemos notar que os diferentes tipos de vestimentas não estão bem delimitados e definidos. Outra forma mais clara de observar esse fenômeno é através do gráfico de dispersão, nele as classes também não estão delimitadas, nos fazendo inferir que o modelo com espaço latente igual a 2 não consegue capturar propriamente a complexidade do dado.

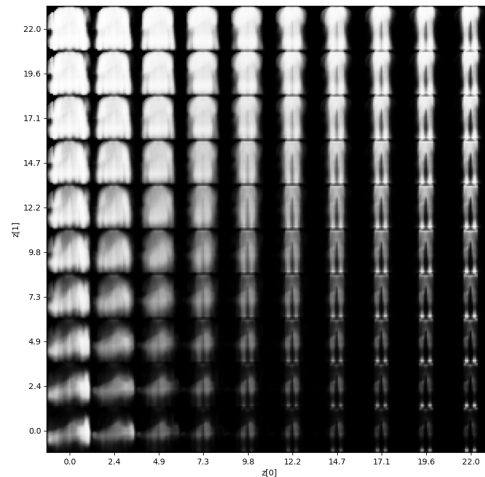
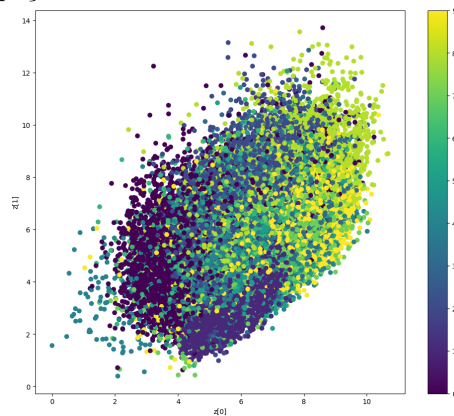


Figure 5: espaço latente do autoencoder de 2 dimensões latentes



### 3 Variational Autoencoder (VAE)

Para o VAE de dimensão latente igual a 64 o treinamento para todo os dados do fashion MNIST levou 144.19 segundos, um pouco mais de tempo que o Autoencoder básico, visto anteriormente. A loss do VAE é normalmente melhor, o treino foi 288.61 onde a parcela da loss de reconstrução foi de 282.07 e loss de The Kullback Leibler responsável pela regularização do modelo foi de 6.9. Segue hiperparametros:

- Dimensão do espaço latente : 64
- Epochs : 30
- Otimizador : Adam
- batch size : 128

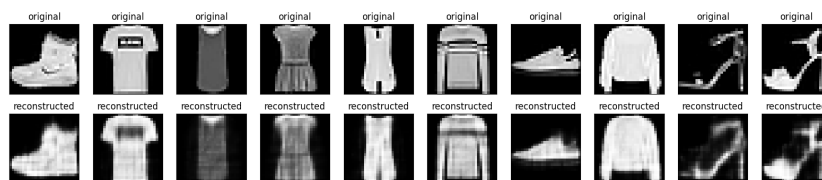


Figure 6: variational autoencoder com 64 dimensões latentes nos dados de treino

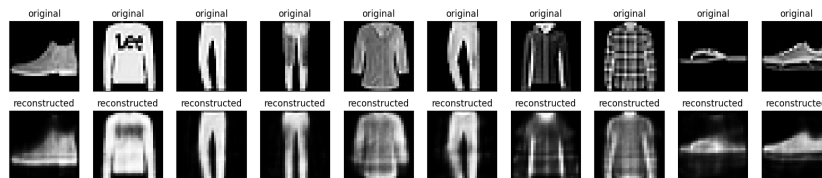


Figure 7: variational autoencoder com 64 dimensões latentes nos dados de teste

Apesar de exigir mais tempo de treinamento, demonstrou melhor capacidade de aprendizado e representação dos dados, com uma loss de reconstrução significativamente menor. Além disso sua matriz de reconstrução apresenta diferentes tipos de vestimentas melhor delimitadas e definidas comparando com o modelo anterior. Provavelmente como resultado do fato do VAE incorporar uma interpretação probabilística ao espaço latente, o que o torna mais adequado para a tarefa.

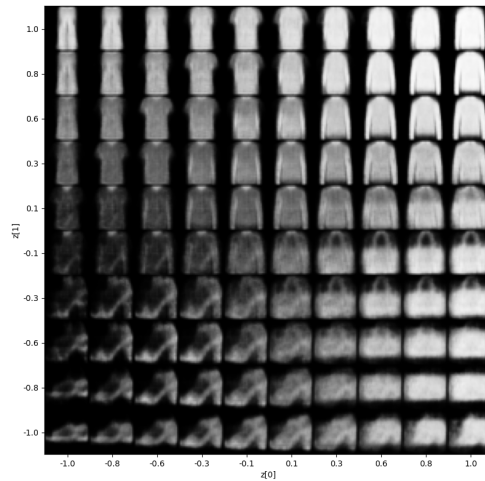


Figure 8: variational autoencoder com 64 dimensões latentes nos dados de treino

## 4 Generative Adversarial Networks (GANs)

Para gerar amostras a partir dos dados do Fashion MNIST foram utilizados os seguintes hiperparâmetros:

- Dimensão do espaço latente : 64
- Epochs : 30
- Otimizador : Adam
- batch size : 256
- distribuição

Note que o otimizador Adam foi utilizado tanto para atualizar os pesos do gerador quanto do discriminador da GAN. Para avaliar seu desempenho geramos imagens utilizando as distribuições normal e uniforme para gerar imagens sintéticas, conforme mostrado na figura a seguir. O treinamento da GAN utilizando a distribuição normal levou 890 segundos e se mostrou promissor em recriar imagens que representem o domínio dos dados de vestimenta. Já para a GAN utilizando a distribuição uniforme levou 830 e se mostrou inferior em criar imagens sintéticas nos parametros apresentados.

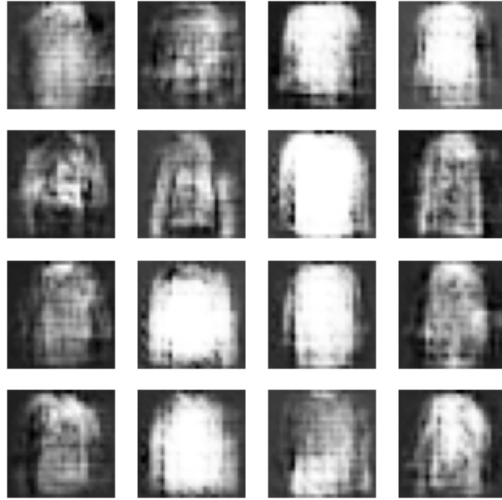


Figure 9: GAN com 64 dimensões latentes e distribuição normal

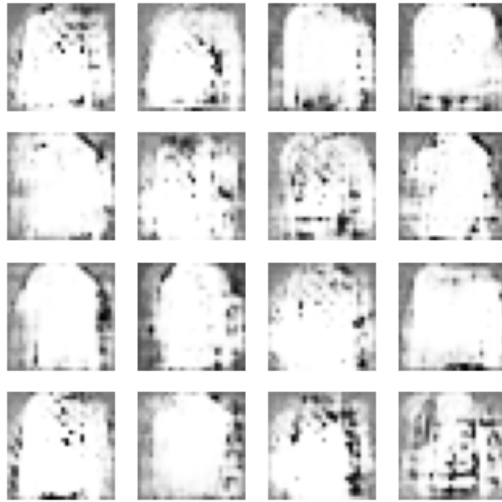


Figure 10: GAN com 64 dimensões latentes e distribuição uniforme

Em conclusão aos modelos apresentados, o Autoencoder apresentou resultados limitados, com o modelo de baixa dimensionalidade de espaço latente (2 dimensões) sendo incapaz de capturar a complexidade dos dados, resultando em reconstruções inadequadas. No entanto, ao aumentar a dimensionalidade do espaço latente para 64, o modelo mostrou uma melhora significativa na capacidade de representação, evidenciando a necessidade de um espaço latente maior para capturar as características intrínsecas dos itens de moda.

Por outro lado, o Variational Autoencoder (VAE), apesar de exigir um tempo de treinamento mais longo, destacou-se com uma capacidade aprimorada de aprendizado e representação dos dados. A abordagem probabilística do VAE no espaço latente contribuiu para uma reconstrução mais precisa e melhor delimitação das diferentes classes de vestuário.

Finalmente, as Generative Adversarial Networks (GANs) apresentaram resultados promissores, sendo capazes de gerar imagens sintéticas que representam o domínio dos dados de vestimenta, especialmente quando alimentadas com uma distribuição normal. No entanto, a utilização da distribuição uniforme resultou em imagens sintéticas inferiores nos parâmetros apresentados.

No geral, cada modelo possui suas próprias vantagens e desvantagens. O Autoencoder é simples e rápido, mas requer um espaço latente maior para desempenho satisfatório. O VAE incorpora a incerteza no espaço latente, tornando-o adequado para tarefas de síntese. As GANs têm o potencial de gerar amostras realistas, mas requerem ajustes cuidadosos e seleção de distribuições adequadas.

Aqui, segue o [link do Colab](#).