

Relatório: Autoencoder e Variational Autoencoder

Almir Sérgio Ramos dos Santos Filho

NES – Redes Neurais

Este relatório apresenta uma análise comparativa entre um Autoencoder clássico e um Variational Autoencoder (VAE), aplicados ao dataset Fashion-MNIST. O objetivo foi observar diferenças entre reconstrução, estruturação do espaço latente e capacidade de geração. As arquiteturas foram mantidas densas para evidenciar os conceitos principais sem depender de otimizações espaciais.

O dataset foi normalizado para o intervalo $[0,1]$ e achatado em vetores de 784 dimensões. O Autoencoder foi estruturado de forma simétrica, reduzindo a entrada para um espaço latente de 32 dimensões. Essa escolha buscou preservar informação suficiente para reconstrução sem tornar o latente muito amplo. O modelo convergiu rapidamente e produziu reconstruções nítidas, especialmente em itens com contornos fortes. Esse bom desempenho ocorre porque o AE não impõe restrições ao formato da representação latente, favorecendo reconstruções fiéis, mesmo que o espaço comprimido não tenha organização interpretável.

O VAE, por outro lado, foi projetado para aprender uma distribuição contínua no espaço latente. O encoder gera vetores de média e variância, e o truque da reparametrização permite amostrar pontos dessa distribuição durante o treino. A perda combina reconstrução e divergência KL, que força o espaço latente a se aproximar de uma normal padrão. Optou-se por duas dimensões para possibilitar visualização direta da organização das classes.

As reconstruções do VAE foram mais suaves que as do AE, o que já era esperado, pois o modelo precisa equilibrar fidelidade e regularização estatística. Entretanto, o mapa 2D do latente mostrou separação clara entre classes e transições contínuas entre categorias semelhantes, indicando que o VAE capturou relações semânticas relevantes. Além disso, a amostragem no espaço latente permitiu gerar novas imagens coerentes, demonstrando a principal vantagem do VAE sobre o Autoencoder tradicional.

A comparação entre os dois modelos evidencia um trade-off: o AE produz melhores reconstruções, enquanto o VAE oferece um espaço latente estruturado e útil para geração e interpolação. As métricas numéricas possuem escalas distintas, então a interpretação visual foi mais adequada neste experimento. Mesmo assim, ambos atingiram convergência estável conforme observado durante o treinamento.

Algumas limitações foram identificadas. O uso exclusivo de camadas densas reduz a qualidade das reconstruções, já que essas camadas não exploram a estrutura espacial da imagem. O espaço latente bidimensional do VAE, embora útil para visualização, restringe a riqueza das representações. Além disso, o número limitado de épocas pode ter impedido refinamentos adicionais.

Para trabalhos futuros, recomenda-se migrar para arquiteturas convolucionais, aumentar a dimensionalidade do espaço latente e testar variantes como β -VAE para melhor separação semântica. Também seria interessante aplicar métricas perceptuais, como SSIM e FID, ou explorar modelos condicionais e híbridos como VAE-GAN para melhorar nitidez e realismo.

Conclui-se que ambos os modelos apresentam vantagens específicas: o Autoencoder é mais eficiente na reconstrução, enquanto o VAE oferece maior capacidade de geração e uma representação latente organizada. Mesmo com arquiteturas simples, o experimento demonstrou de forma clara os princípios fundamentais que diferenciam essas duas abordagens dentro da família de modelos gerativos.