

## EP 4

Data de entrega: 30/11/2024 (como sempre, negociável!)

O objetivo deste EP será experimentar arquiteturas do tipo *autoencoder*, listadas a seguir.

**Auto-encoder:** é uma arquitetura típica usada principalmente para redução de dimensão dos dados. A estrutura é do tipo *encoder-decoder*, tendo a particularidade de a entrada e a saída *target* serem as mesmas. Por isso é muitas vezes considerada uma rede para redução de dimensionalidade ou compressão de dados. O que se tenta otimizar é a reconstrução do dado de entrada.

Quando utilizado com imagens, a etapa de codificação reduz progressivamente as dimensões da imagem, comprimindo-a em uma representação de menor dimensão que preserva as características essenciais. De forma análoga, a etapa de decodificação parte dessa representação comprimida para reconstruir a imagem de saída com a maior precisão possível.

**Denoising auto-encoder:** é basicamente um *auto-encoder*, com a diferença de que os dados de entrada são ligeiramente perturbados com a adição de ruídos aleatórios. Além de verificar que a rede “aprende” a filtrar ruídos, vale a pena verificar se existe alguma diferença em relação ao *auto-encoder* anterior, quando avaliamos o desempenho deste em relação a entradas sem perturbação.

**Variational auto-encoder (VAE):** Também segue o mesmo princípio do *auto-encoder* em termos de entrada e saída, porém o principal foco neste modelo é fazer com que o espaço latente (a representação comprimida gerada entre a parte *encoder* e *decoder*) esteja associado a uma distribuição de probabilidade. Com isso, pequenas transições no espaço latente deveriam refletir pequenas variações nas correspondentes reconstruções, dando abertura para a geração de dados “novos” similares aos usados no treinamento.

Note que enquanto as duas primeiras redes tem como motivação principal a redução de dimensão ou mesmo a extração de *features*, a terceira rede tem como motivação a geração de novos dados (similares às entradas usadas no treinamento). Em todo caso, essas redes são comparáveis entre elas com respeito a algum aspecto? (não é necessário responder essa pergunta, apenas refletir sobre ela)

## O que fazer?

Você deve implementar e experimentar as duas primeiras redes. Existem várias arquiteturas de autoencoders, sendo o Convolutional Autoencoder uma excelente opção para iniciar a experimentação.

Para a terceira rede, uma versão inicial do código será disponibilizada em breve e você poderá experimentar partindo desse código. Caso prefira, sintase à vontade para fazer a sua própria implementação.

Como das vezes anteriores, para as implementações um código já existente pode ser usado como ponto de partida. Se for o caso, indique de onde ele foi retirado, se e quais adaptações você fez, etc

Quanto aos dados a serem usados, uma sugestão é o MNIST ou algum conjunto de imagens não muito grandes nem muito complexos<sup>1</sup>. Imagens vêm acompanhadas da possibilidade de inspeção visual, mais qualitativa, do resultado – um motivador extra! Caso queira aplicar sobre outro tipo de dado (que não imagem), fique à vontade.

Em sua experimentação, deverá pensar ao menos um aspecto a ser explorado (uma pergunta sua) e organizar o notebook/script de forma a mostrar claramente o percurso dessa exploração e os resultados obtidos (juntamente com comentários pertinentes). Fiquem à vontade para usar técnicas de visualização que ajudam a entender melhor os dados, resultados, *features*, diferenças, etc.

## Exploração adicional

Como de praxe, durante a realização desta tarefa, é esperado que você desenvolva uma postura crítica e investigativa, formulando suas próprias perguntas e trabalhando as dúvidas sobre o exercício proposto.

A seguir, alguns exemplos de perguntas para guiar a exploração.

- Qual é o impacto de aumentar ou reduzir a dimensão da representação intermediária?
- O t-SNE é uma técnica popular para visualizar dados de alta dimensão em um espaço de duas ou três dimensões, permitindo observar agrupamentos no dataset. Você consegue ver a diferença entre a visualização do dataset original e da visualização da representação comprimida (saída do encoder) do mesmo dataset?

---

<sup>1</sup>Cuidado com a questão do custo computacional!