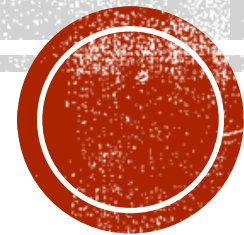


REDES NEURAIS ARTIFICIAIS

AUTOENCODERS

Prof. Rodrigo Palácios

rodrigopalacios@utfpr.edu.br



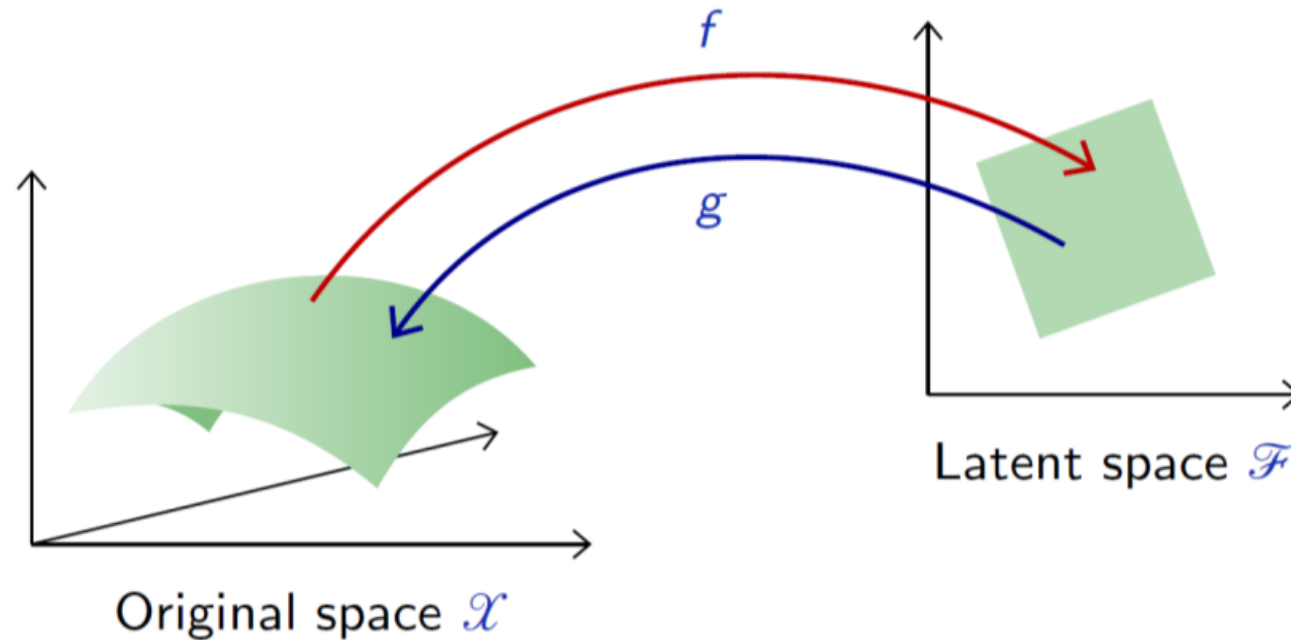
AUTOENCODERS – INTRODUÇÃO

- Os Autoencoders são uma técnica de aprendizado não supervisionado
 - Usa-se redes neurais para a tarefa de aprendizado de representação.
 - Projeta-se uma arquitetura de rede neural de modo a impor um gargalo na rede que força uma representação de conhecimento compactada da entrada original.
 - Se os recursos de entrada fossem independentes um do outro, essa compressão e reconstrução subsequente seriam uma tarefa muito difícil. No entanto, se houver algum tipo de estrutura nos dados (ou seja, correlações entre os recursos de entrada), essa estrutura poderá ser aprendida e consequentemente aproveitada ao forçar a entrada através do gargalo da rede.
- Autoencoders (AE) são redes neurais que visam copiar suas entradas para suas saídas. Eles trabalham compactando a entrada em uma representação de espaço latente e, em seguida, reconstruindo a saída dessa representação. Esse tipo de rede é composto de duas partes:
 - Codificador (Encoder): é a parte da rede que compacta a entrada em uma representação de espaço latente (codificando a entrada). Pode ser representado por uma função de codificação $h = f(x)$.
 - Decodificador (Decoder): Esta parte tem como objetivo reconstruir a entrada da representação do espaço latente. Pode ser representado por uma função de decodificação $r = g(h)$.

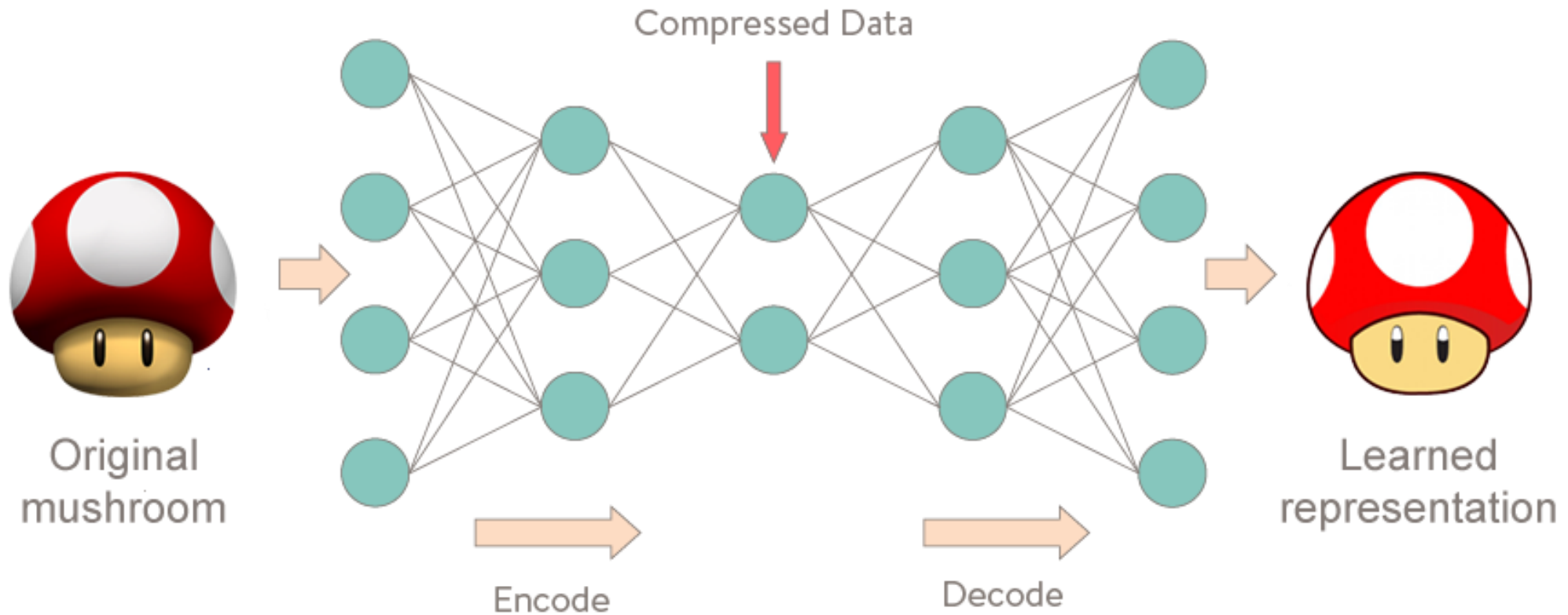
AUTOENCODER – INTRODUÇÃO

- Autoencoder combina um codificador f do espaço original X para um espaço latente F , e um decodificador g mapear de volta para X , tal que $f \circ g$ está próximo da identidade nos dados

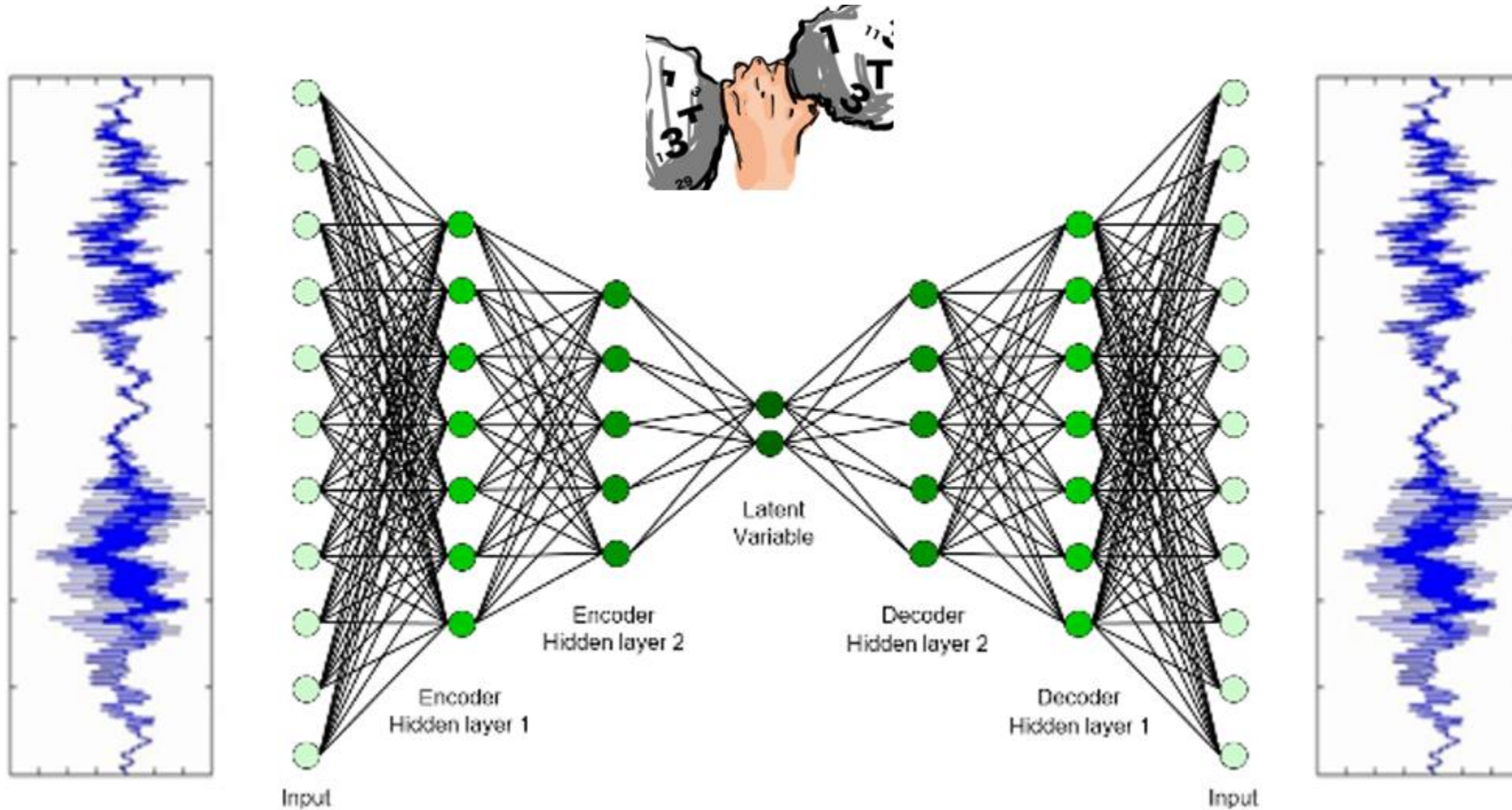
$$E[\|X - g \circ f(X)\|^2] \approx 0$$



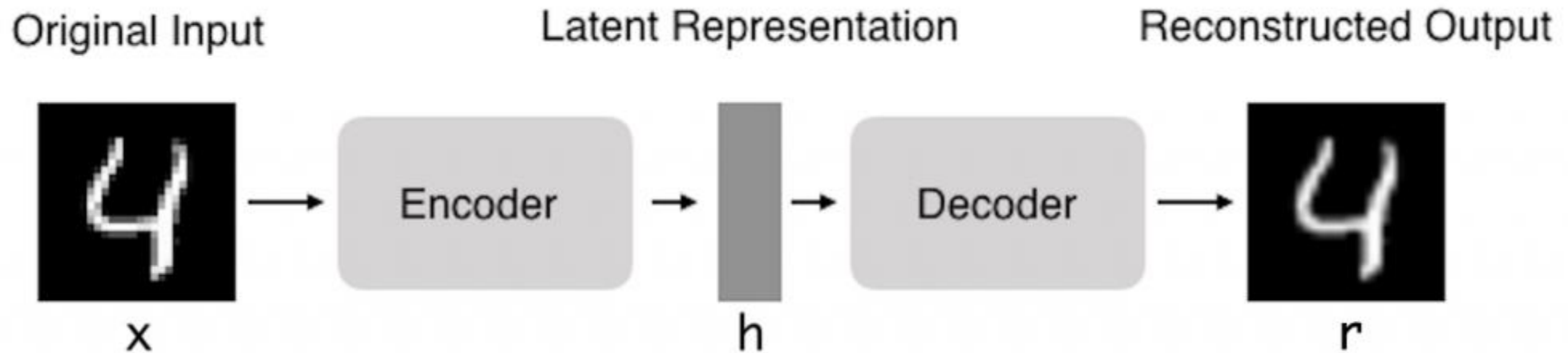
AUTOENCODERS – INTRODUÇÃO



AUTOENCODER – INTRODUÇÃO



AUTOENCODERS – INTRODUÇÃO



AUTOENCODERS – INTRODUÇÃO

- A função dos Autoencoders não é apenas copiar a entrada para a saída.
- No uso, é esperado que treinando o Autoencoder para copiar a entrada para a saída a representação latente h tenha propriedades úteis.
 - Pode ser conseguido criando restrições na tarefa de cópia.
 - Uma maneira de obter recursos úteis do Autoencoder é restringir h a ter dimensões menores que x , nesse caso, o Autoencoder é chamado de incompleto.
 - Ao treinar uma representação incompleta, forçamos o Autoencoder a aprender os recursos mais importantes dos dados de treinamento.
 - Se for dada muita capacidade ao Autoencoder, ele poderá aprender a executar a tarefa de cópia sem extrair nenhuma informação útil sobre a distribuição dos dados. Isso também pode ocorrer se a dimensão da representação latente for a mesma que a entrada e, no caso de excesso de conclusão, em que a dimensão da representação latente for maior que a entrada.

AUTOENCODERS – APLICAÇÕES

- As aplicações mais tradicionais dos Autoencoders são a remoção de ruídos e a redução de dimensionalidade para visualização de dados.
- Com restrições de dimensionalidade e esparsidade apropriadas, os Autoencoders podem aprender projeções de dados mais interessantes.
- Os Autoencoders aprendem automaticamente a partir de exemplos de dados. Isso significa que é fácil treinar instâncias especializadas do algoritmo que terão bom desempenho em um tipo específico de entrada e que não requer nenhuma nova engenharia de recursos, apenas os dados de treinamento apropriados.
- Outra aplicação interessante dos Autoencoders é como tarefa preliminar ao reconhecimento de imagens com CNNs (Redes Neurais Convolucionais), por exemplo.
- Os Autoencoders são treinados para preservar o máximo de informações possível quando uma entrada é passada pelo codificador e depois pelo decodificador, mas também são treinados para fazer com que a nova representação tenha várias propriedades agradáveis.
- Diferentes tipos de Autoencoders visam atingir diferentes tipos de propriedades.

TIPOS DE REDES NEURAIS ARTIFICIAIS AUTOENCODERS

▪ Autoencoder Padrão

- Na sua forma mais simples, o Autoencoder é uma rede neural artificial de três camadas, isto é, uma rede neural com uma camada de entrada, uma oculta e uma camada de saída.
- A entrada e a saída são as mesmas e aprendemos a reconstruir a entrada, por exemplo, usa-se o otimizador adam e a função de perda de erro quadrático médio.

▪ Autoencoder Multicamada

- É possível estender o Autoencoder para mais camadas ocultas, caso uma única camada não seja suficiente.
- Qualquer uma das camadas ocultas pode ser escolhida como representação de recurso, mas o ideal é tornar a rede simétrica e usar a camada mais intermediária.

▪ Autoencoder Convolucional

- Os Autoencoders podem ser usados com convoluções em vez de camadas totalmente conectadas.
- O princípio é o mesmo, mas usando imagens (vetores 3D) em vez de vetores 1D achatados (flattened).
- A imagem de entrada é reduzida para fornecer uma representação latente de dimensões menores e forçar o Autoencoder a aprender uma versão compactada das imagens.

TIPOS DE REDES NEURAIS ARTIFICIAIS AUTOENCODERS

- **Autoencoder Regularizado**

- Existem outras maneiras pelas quais podemos restringir a reconstrução de um Autoencoder, além de impor uma camada oculta de menor dimensão que a entrada. Em vez de limitar a capacidade do modelo mantendo o codificador e o decodificador rasos e o tamanho do código pequeno, os Autoencoders regularizados usam uma função de perda que incentiva o modelo a ter outras propriedades além da capacidade de copiar sua entrada para sua saída.
- **Autoencoder Esparso:** geralmente são usados para aprender recursos para outra tarefa, como classificação. Um Autoencoder que foi regularizado para ser esparsos deve responder a recursos estatísticos exclusivos do conjunto de dados em que foi treinado, em vez de simplesmente atuar como uma função de identidade. Dessa forma, o treinamento para executar a tarefa de cópia com uma penalidade de escassez pode produzir um modelo que aprendeu recursos úteis como subproduto.
- **Autoencoder Denoising:** em vez de adicionar uma penalidade à função de perda, podemos obter um Autoencoder que aprende algo útil alterando o termo do erro de reconstrução da função de perda. Isso pode ser feito adicionando algum ruído à imagem de entrada e fazendo o Autoencoder aprender a removê-la. Dessa maneira, o Autoencoder extrairá os recursos mais importantes e aprenderá uma representação robusta dos dados.

TIPOS DE REDES NEURAIS ARTIFICIAIS AUTOENCODERS

- **Contractive Autoencoders (CAE)**

- O objetivo do Autoencoder Contrativo (CAE) é ter uma representação aprendida robusta, menos sensível a pequenas variações nos dados. A robustez da representação para os dados é feita aplicando um termo de penalidade à função de perda. É outra técnica de regularização, como os Autoencoders Esparsos e os Autoencoders Denoising.

- **Deep Autoencoders**

- Deep Autoencoders consistem em duas redes de crenças profundas idênticas (Deep Belief Networks). Uma rede para codificação e outra para decodificação. Os Autoencoders tipicamente profundos têm de 4 a 5 camadas para codificação e as próximas 4 a 5 camadas para decodificação. Usa-se camada não supervisionada por camada, pré-treinamento.

- **Variational Autoencoders (VAEs)**

- É um Autoencoder cuja distribuição de codificações é regularizada durante o treinamento, a fim de garantir que seu espaço latente tenha boas propriedades, o que nos permite gerar novos dados. Além disso, o termo “variacional” vem da estreita relação que existe entre a regularização e o método de inferência variacional em Estatística.

REFERÊNCIAS

- <http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/>
- <https://www.jeremyjordan.me/autoencoders/>
- <https://www.deeplearningbook.com.br/introducao-aos-autoencoders/>