Autoencoder

Dr. Alireza Aghamohammadi

We Should Learn Unsupervised Learning First

❖ Before discussing autoencoders, let's explore unsupervised learning, a type of learning algorithm.

❖ In unsupervised learning, we only have input data x without any labels.

❖ The objective is to uncover the **hidden** or **underlying structure** within the data.

❖ Examples of unsupervised learning include:
   ❑ Feature extraction
   ❑ Dimensionality reduction
   ❑ Outlier detection

What Is Autoencoder?

- ❖ Autoencoders learn lower-dimensional feature representations from unlabeled data.
- ❖ **Encoder**: Maps data $x \in \mathbb{R}^m$ to latent space $z \in \mathbb{R}^n$ ($m > n$).
- ❖ Latent space z is not directly observable.
- ❖ Model is trained to reconstruct original data from z.
- ❖ **Decoder**: Maps latent space z back to reconstructed data.

$$\mathcal{L} = \frac{1}{N} \sum \|x - \hat{x}\|^2$$

$x \rightarrow \boxed{} \rightarrow z \rightarrow \boxed{} \rightarrow \hat{x}$