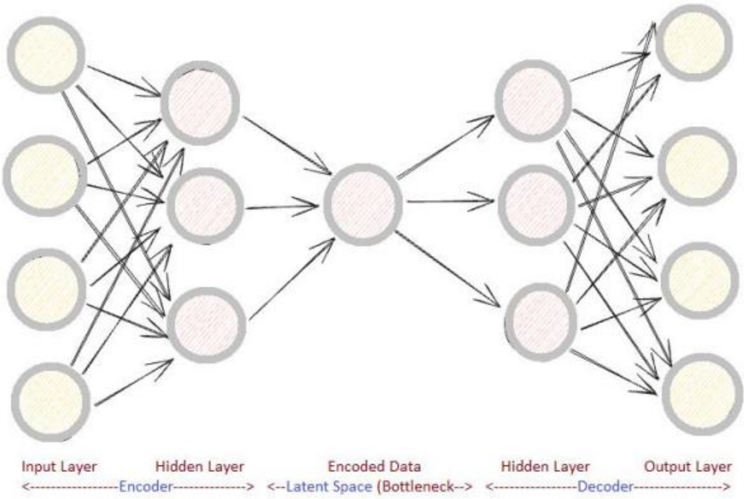
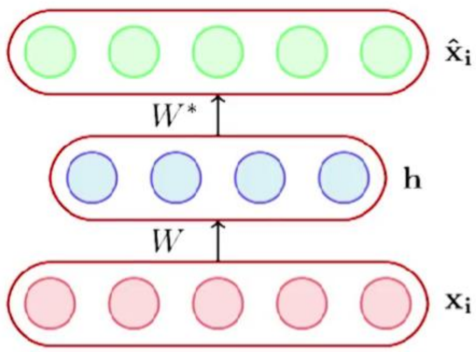
Undercomplete Autoencoder

- An undercomplete autoencoder is one of the simplest types of autoencoders. Undercomplete autoencoder takes in an image and tries to predict the same image as output, thus reconstructing the image from the compressed bottleneck region.
- Undercomplete autoencoders are truly unsupervised as they do not take any form of label, the target being the same as the input.
- The primary use of autoencoders like such is the generation of the latent space or the bottleneck, which forms a compressed substitute of the input data and can be easily decompressed back with the help of the network when needed.
- This form of compression in the data can be modeled as a form of dimensionality reduction.
- In this case, we don't have an explicit regularization mechanism, but we ensure that the size of the bottleneck is always lower than the original input size to avoid overfitting.
- This type of configuration is typically used as a dimensionality reduction technique (more powerful than PCA since its also able to capture non-linearities in the data).
- Under complete autoencoders is an unsupervised neural network that you can use to generate a compressed version of the input data.
- It is done by taking in an image and trying to predict the same image as output, thus reconstructing the image from its compressed bottleneck region.

An autoencoder where $\dim(\mathbf{h}) < \dim(\mathbf{x_i})$ is called an under complete autoencoder



Another example:



$$\mathbf{h} = g(W\mathbf{x}_\mathbf{i} + \mathbf{b})$$
$$\hat{\mathbf{x}}_\mathbf{i} = f(W^*\mathbf{h} + \mathbf{c})$$

- Let us consider the case where $\dim(\mathbf{h}) < \dim(\mathbf{x}_\mathbf{i})$
- If we are still able to reconstruct $\hat{\mathbf{x}}_\mathbf{i}$ perfectly from $\mathbf{h}$, then what does it say about $\mathbf{h}$?
- $\mathbf{h}$ is a loss-free encoding of $\mathbf{x}_\mathbf{i}$. It captures all the important characteristics of $\mathbf{x}_\mathbf{i}$
- Do you see an analogy with PCA?