Guided Tour of Machine Learning in Finance

Week 3: Unsupervised Learning
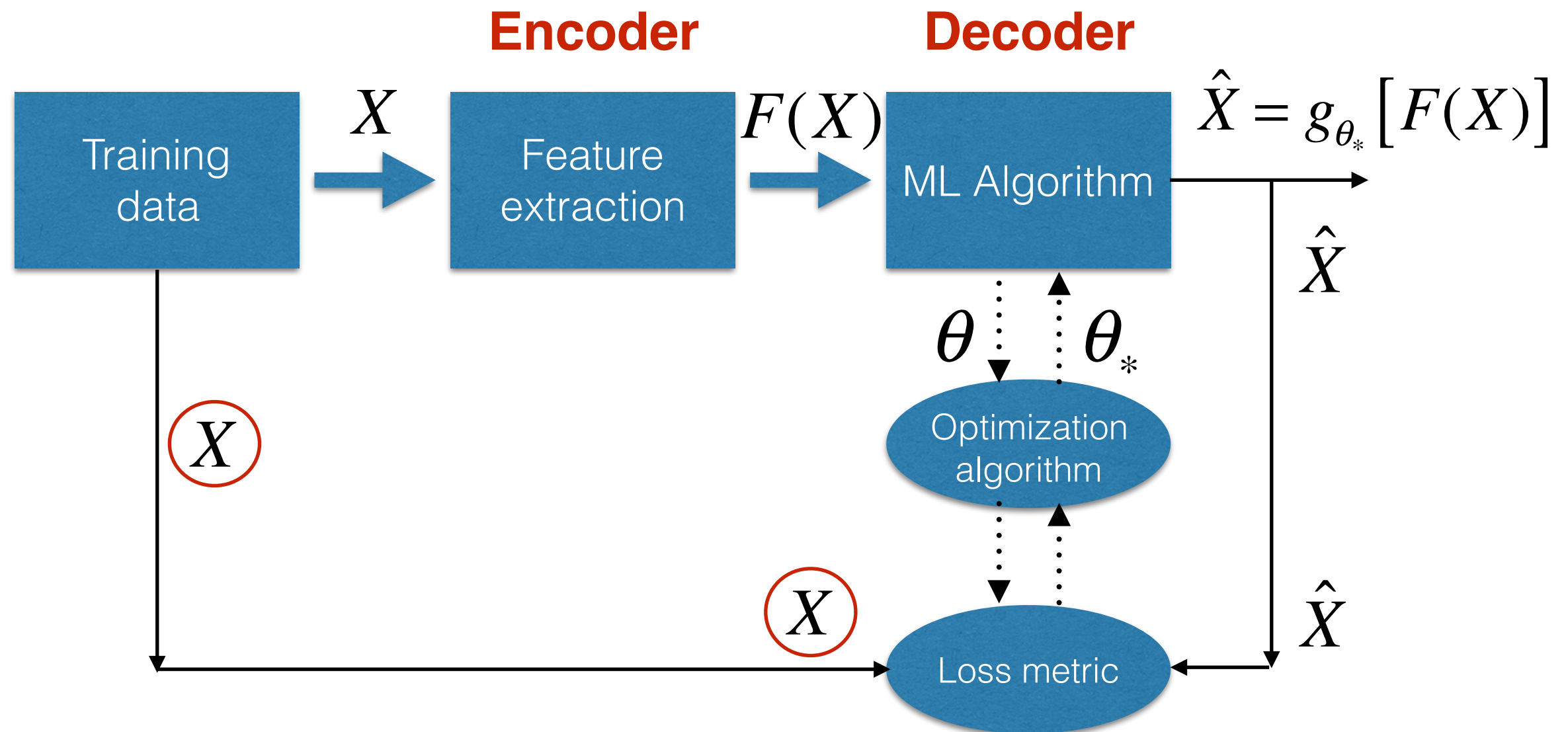
3-1-5-Autoencoder

Igor Halperin

NYU Tandon School of Engineering, 2017

Unsupervised learning diagram: Autoencoder

Autoencoder: replace Y with X!

PCA as a dimension reduction method

We had a linear transform (linear encoder) of the data $\mathbf{Z} = \mathbf{XV}$ parametrized by a $p \times k$ orthogonal matrix $\mathbf{V}$ with $\mathbf{VV}^T = 1$
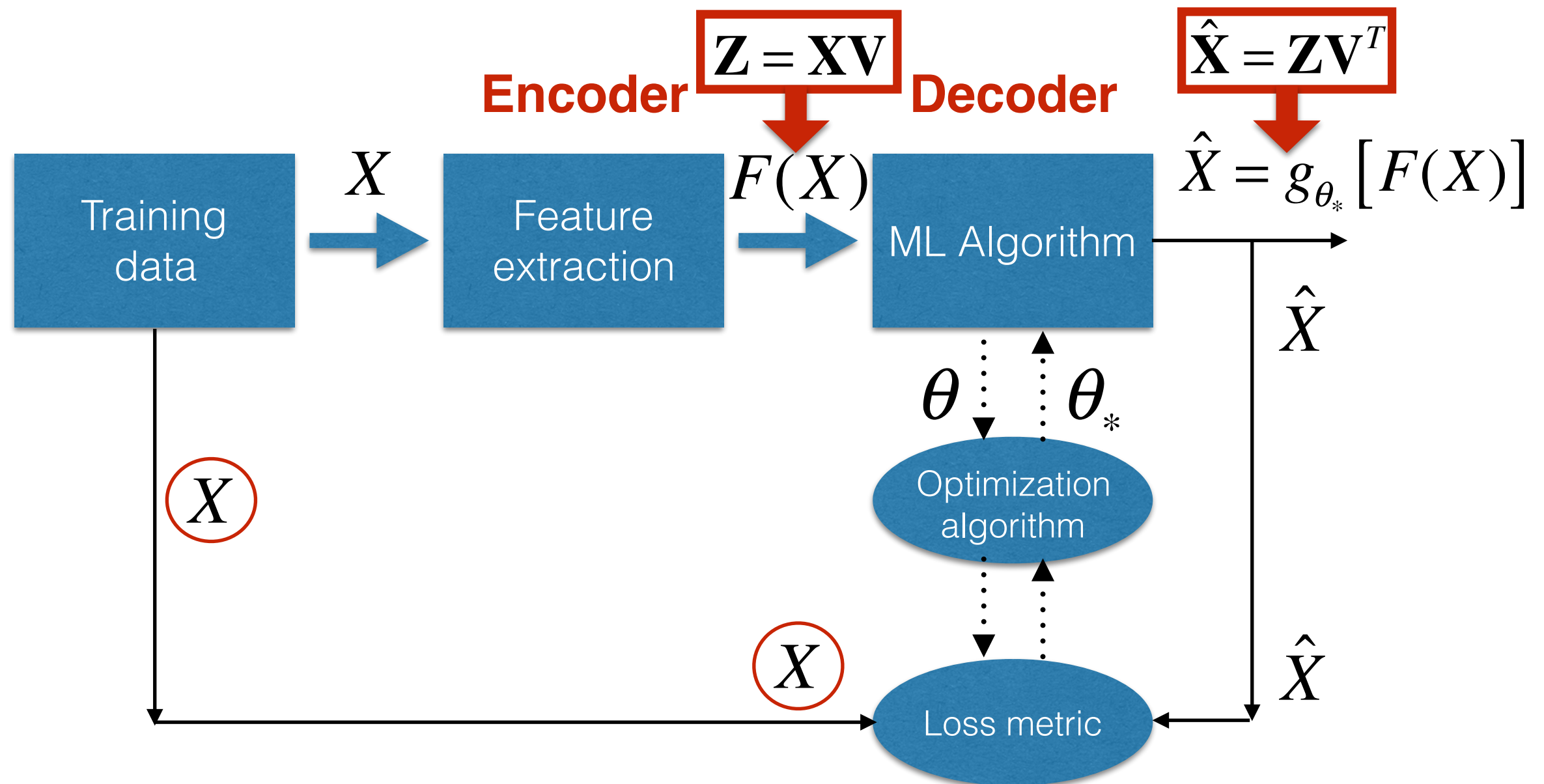

If $k = p$, $\mathbf{V} = \mathbf{U}$ this is just a rotation of the eigenvalue decomposition

A **decoded signal** (<u>inverse transform</u>) is obtained as $\hat{\mathbf{X}} = \mathbf{ZV}^T$
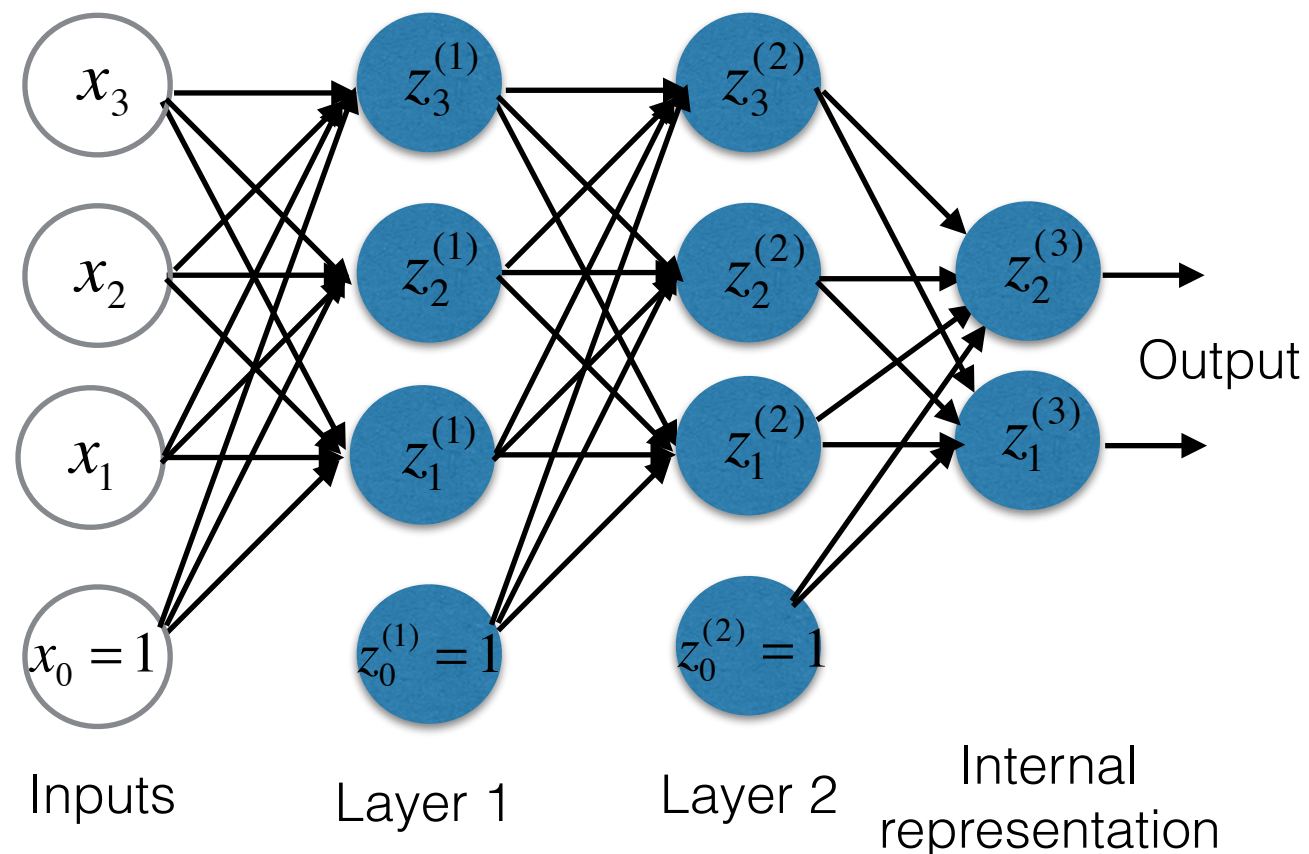
Unsupervised learning diagram: Autoencoder
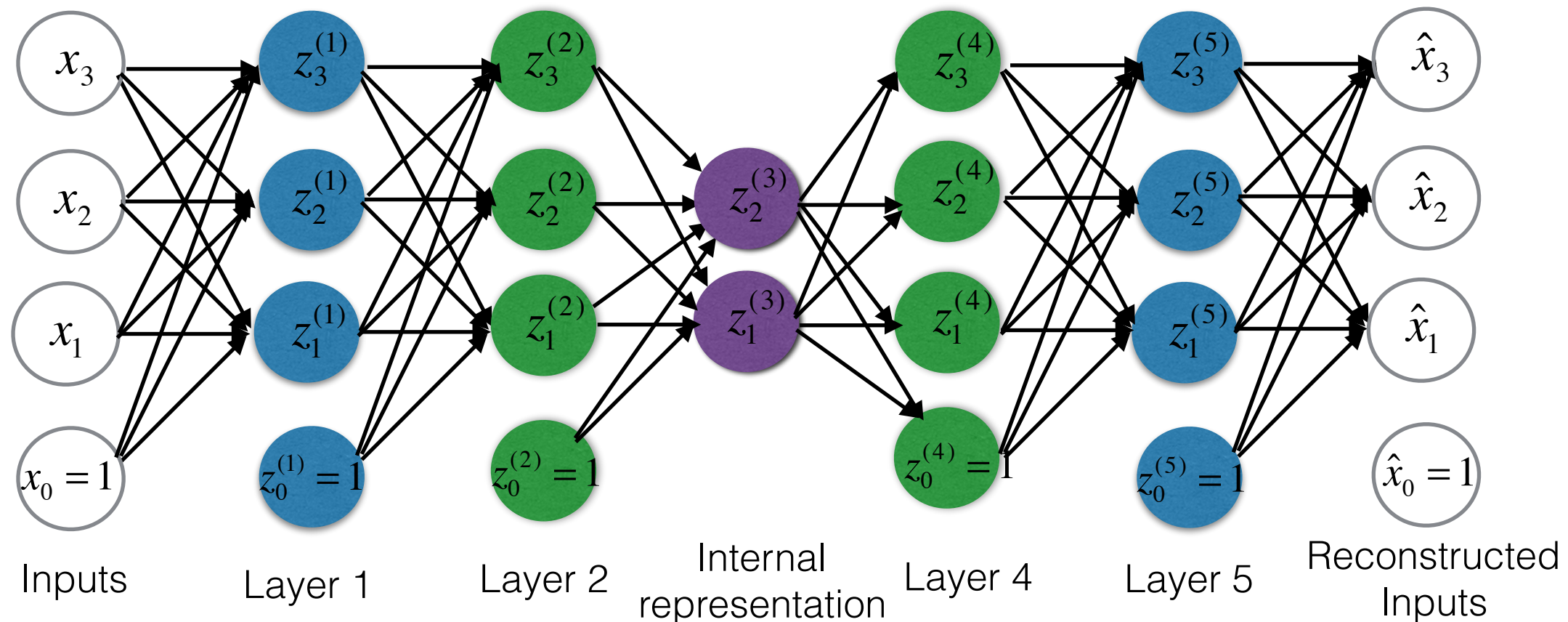
Autoencoder:

Neural Encoder

- **A Neural Encoder** is composed of layers of artificial neurons (perceptrons), with a final layer providing a low-dimensional representation of data

- For linear activation functions in all layers, this is equivalent to a single linear layer, which is equivalent to the Linear Encoder of the PCA
- If we use non-linear activations, then the depth of the network matters. For complex data, many layers might be needed to build a good internal representation.

Neural Encoder-Decoder

- **A Neural Encoder/Decoder pair implements an Autoencoder:**

| Inputs | Layer 1 | Layer 2 | Internal representation | Layer 4 | Layer 5 | Reconstructed Inputs |

- If the inner-most layer has only two neurons, they activation can be plotted for visualization of data with the autoencoder.
- A Stacked (multi-level) autoencoder may have layers with gradually decreasing number of neurons per layer
- Tied weights in the Encoder and Decoder are often used in practice

Visualizing data with autoencoder

Autoencoder provides a way to visualize a high-dimensional input vector by a non-linear projection onto a 2D space of neuron activations for the inner-most layer with 2 neurons:
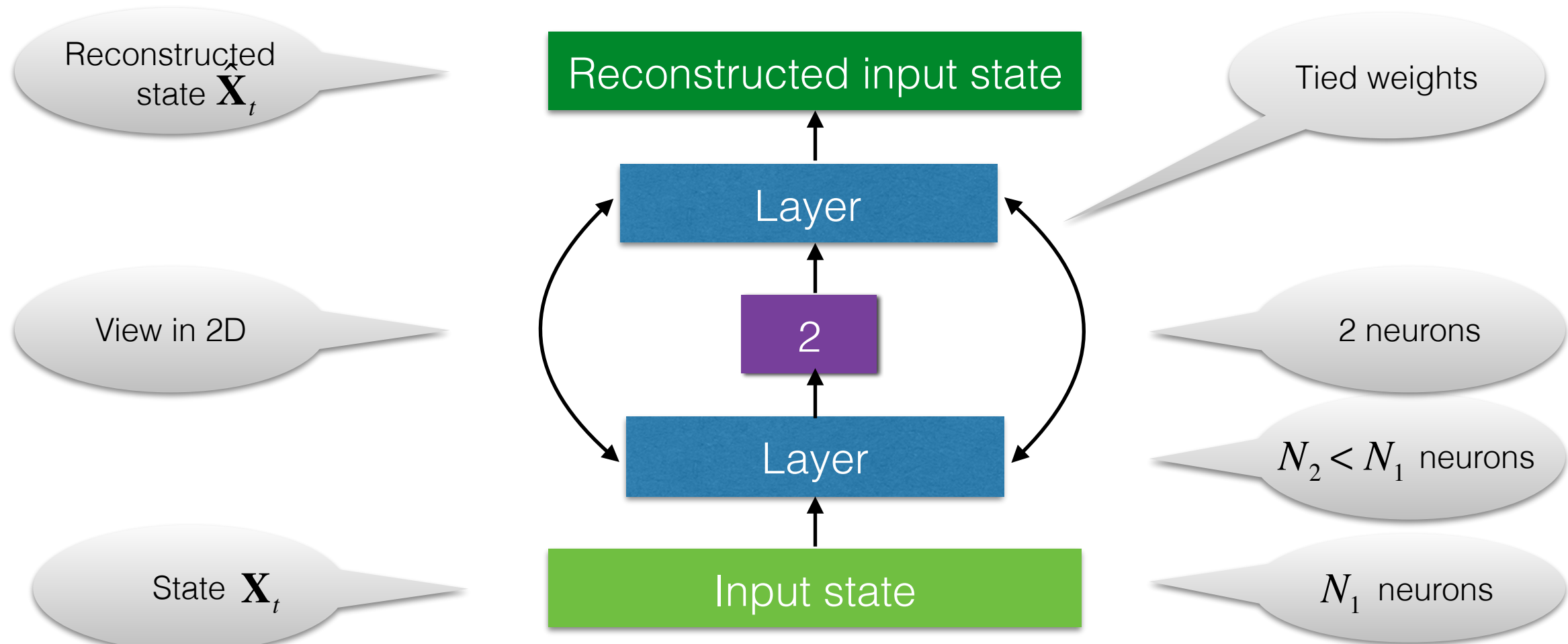
Autoencoder is trained by minimizing the reconstruction error between input state $\mathbf{X}_t$ and reconstructed state $\hat{\mathbf{X}}_t$

Control question

Select all correct answers

1. The PCA implements a Linear Encoder that converts correlated inputs $\mathbf{X}$ into uncorrelated features $\mathbf{Z}$ by a linear transform $\mathbf{Z} = \mathbf{XV}$
2. If the orthogonal matrix $\mathbf{V}$ has dimension $p \times p$, i.e. it keeps all eigenvectors of correlation matrix of $\mathbf{X}$, then $\mathbf{Z}$ preserves the total variation of $\mathbf{X}$.
3. When the data is noisy or non-stationary, $\mathbf{Z}$ can have a higher variance than $\mathbf{X}$.
4. The PCA is a probabilistic method that enables simulating from data.

Correct answer: 1, 2