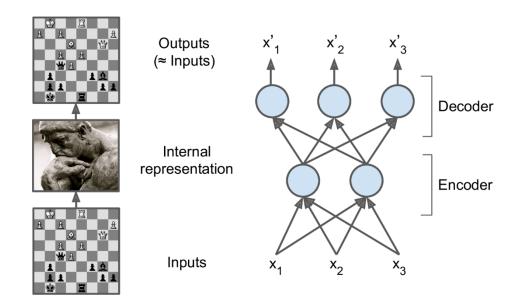
Autoencoders

CMPUT 328

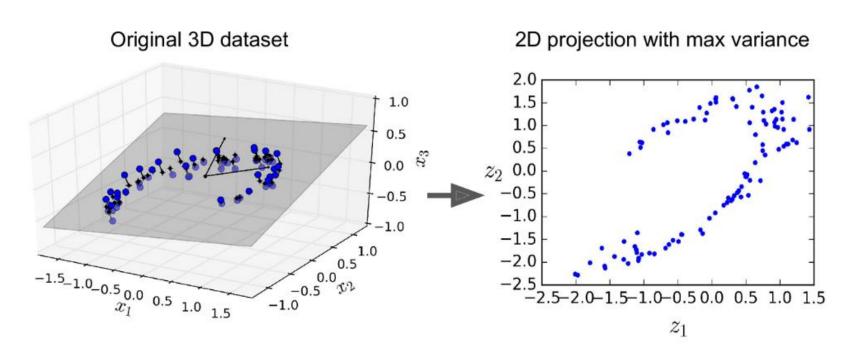
Nilanjan Ray

Representation learning by autoencoders



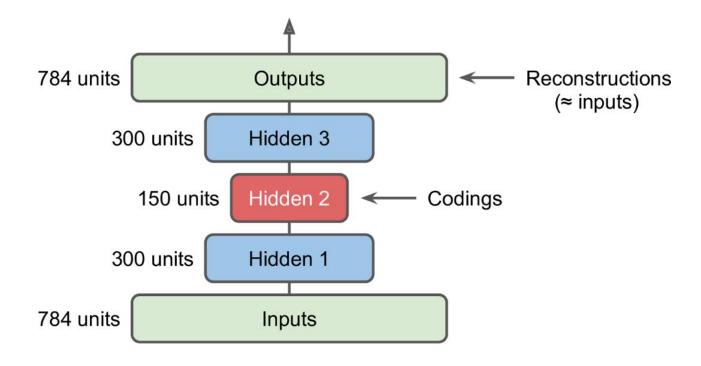
Can we learning interesting, hidden representations from unlabeled data?

Extract underlying (low) dimensionality



Even though the data points are 3D, they more or less lie on a 2D plane.

An example autoencoder for MNIST

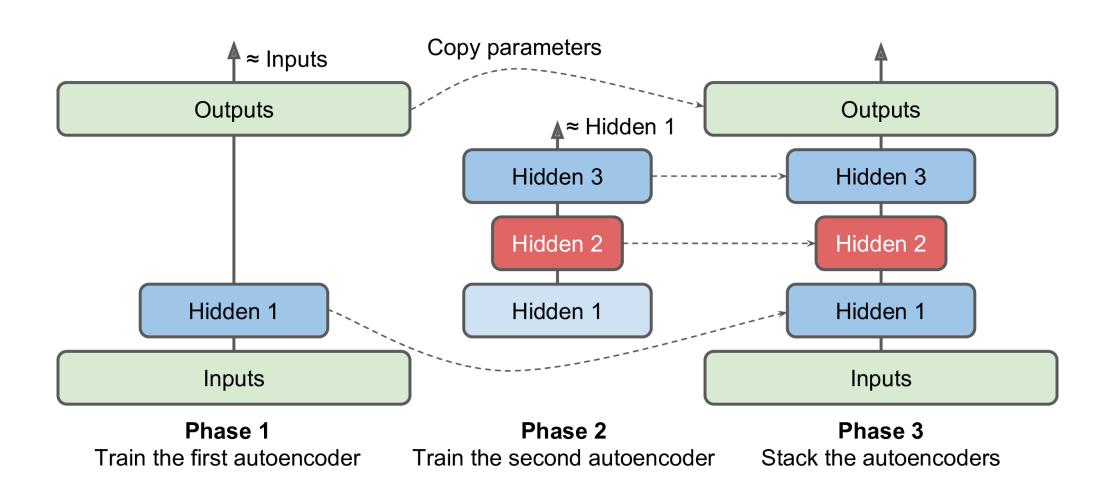


MNIST_AE.ipnyb implements this architecture

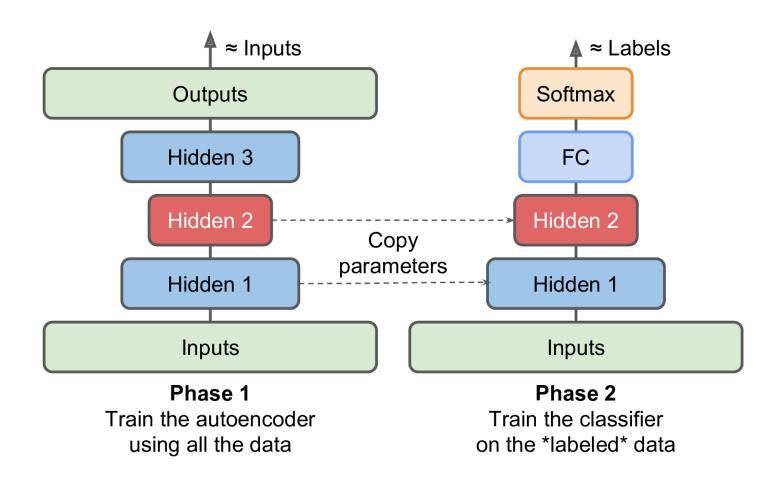
Convolutional AE

- We can easily replace fully connected layers by all convolutional layers with proper padding
- Look at MNIST_AE.ipnyb

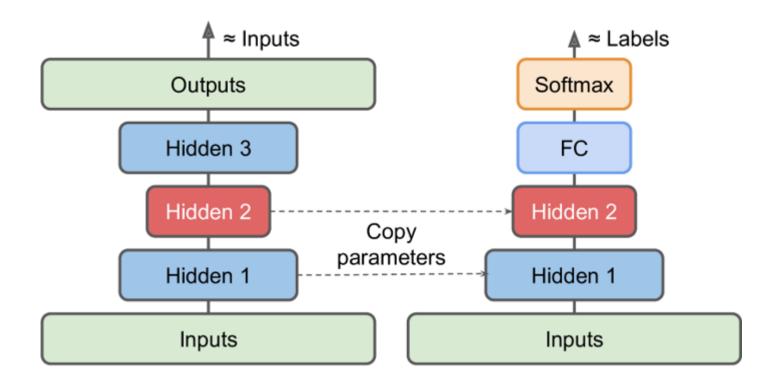
Phased training in AE



Classification from AE

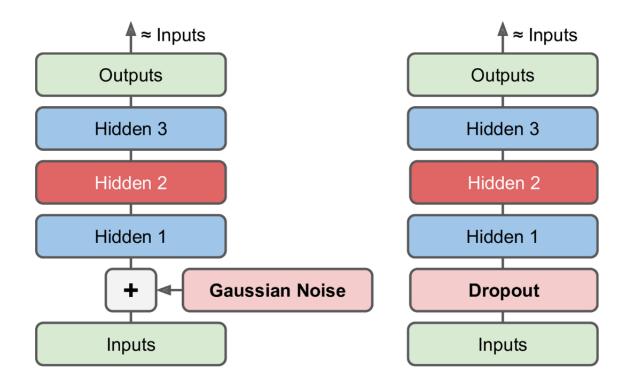


Semi-supervised learning



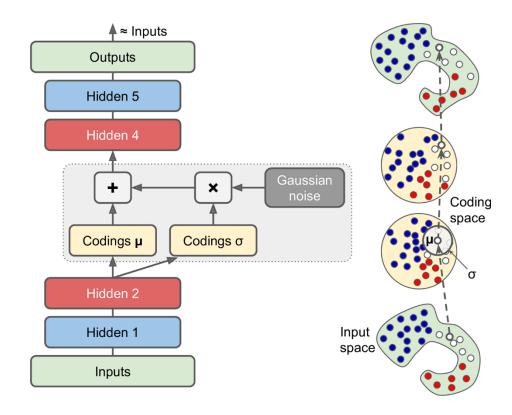
Often, labeled data is partially available: maybe 80% training data is unlabeled, only 20% is labeled. How do we make use of unlabeled data during training a classifier?

Denoising AE



Add noise or insert a dropout layer

Variational AE



Cost function has two components:

Reconstruction + constraint for μ and σ

The constraint: the encoded distribution should look like a zero-mean, unit variance Gaussian.

The advantage is that you can generate data (images) that look like the training images.