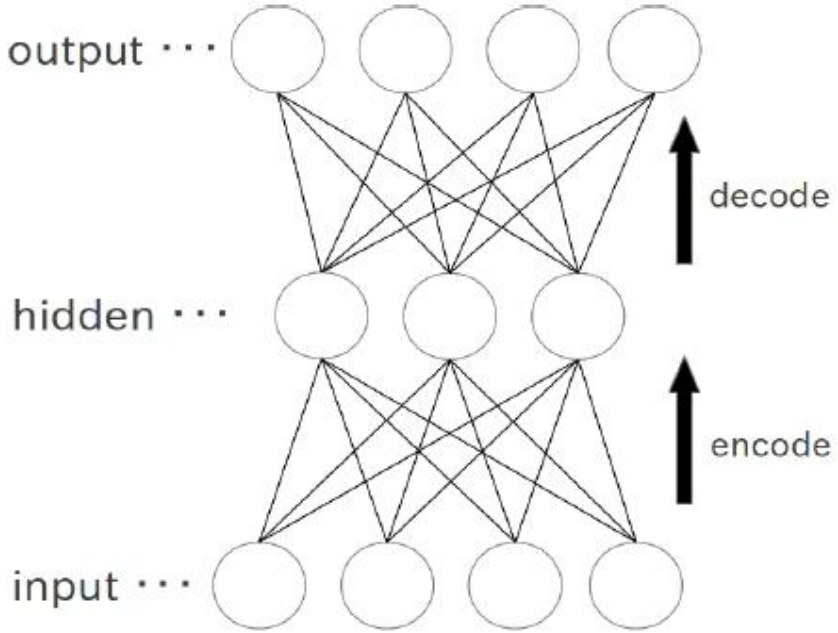
Reducing the Dimensionality of Data with Neural Networks

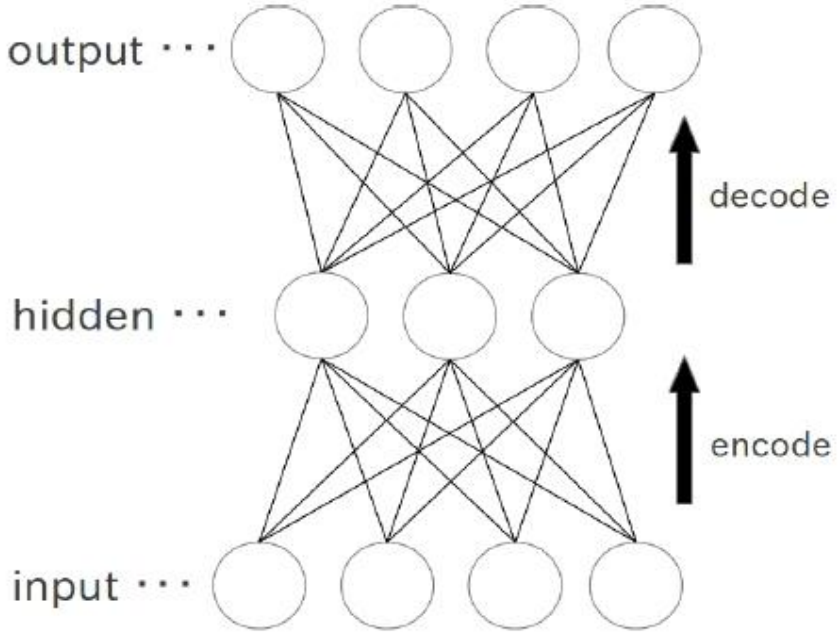
G. E. Hinton* and R. R. Salaktudinov(2006)

Autoencoders



- Hidden Layer lowers dimensionality of input
- Non-linear reduction
- Reconstruct data from the middle layer
- Can perform better than PCA

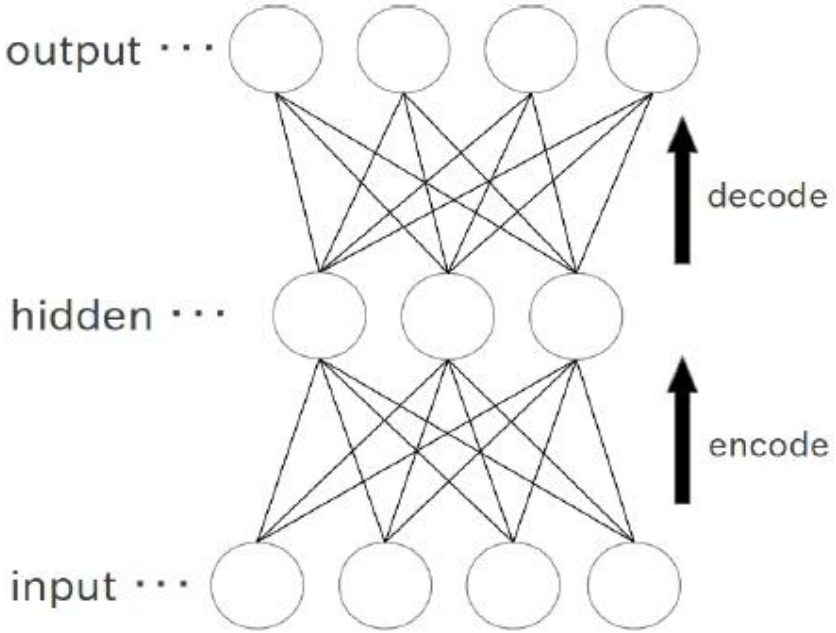
Autoencoders - Problems



Problems

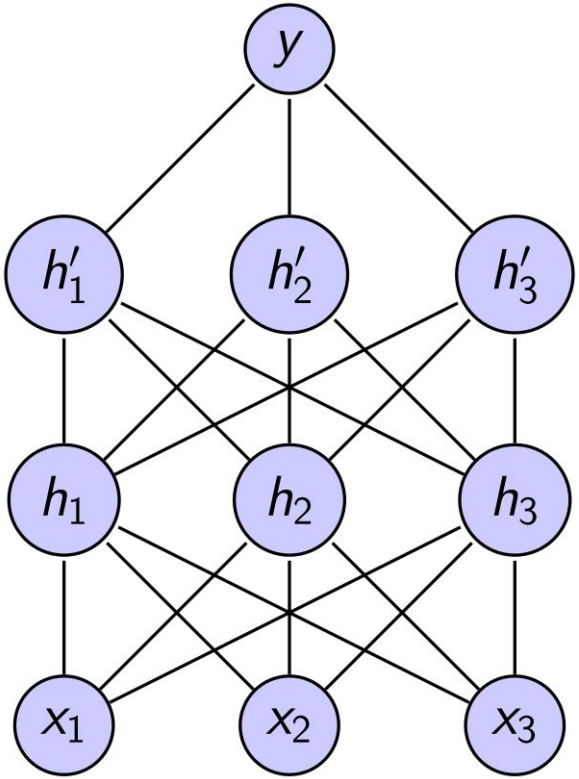
- Vanishing gradient problem for multiple layers
- Bigger NNs requires good initial weights
- How to initialize the weights?

Autoencoders - Pretraining



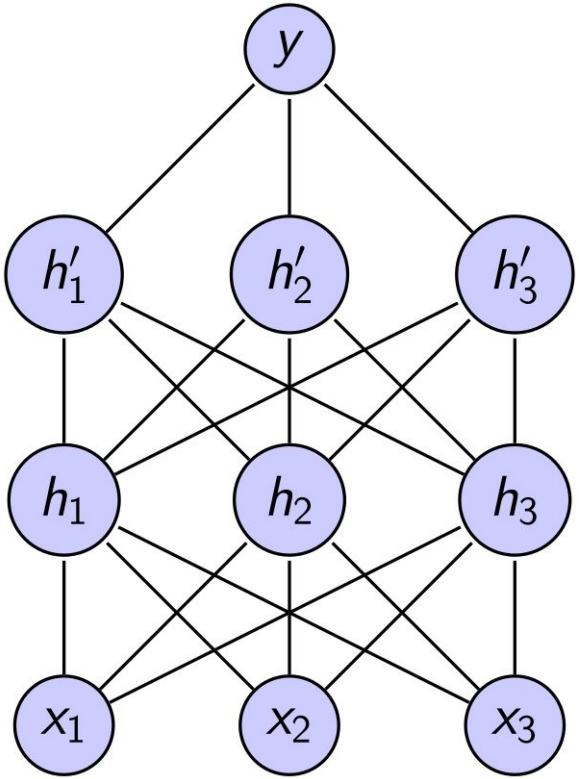
- Train each layer individually in an unsupervised way
- Composed by stacks of **Restricted Boltzmann Machines** (RBM)
- The input vector is binary for explanation purposes

RBMs - Pretraining



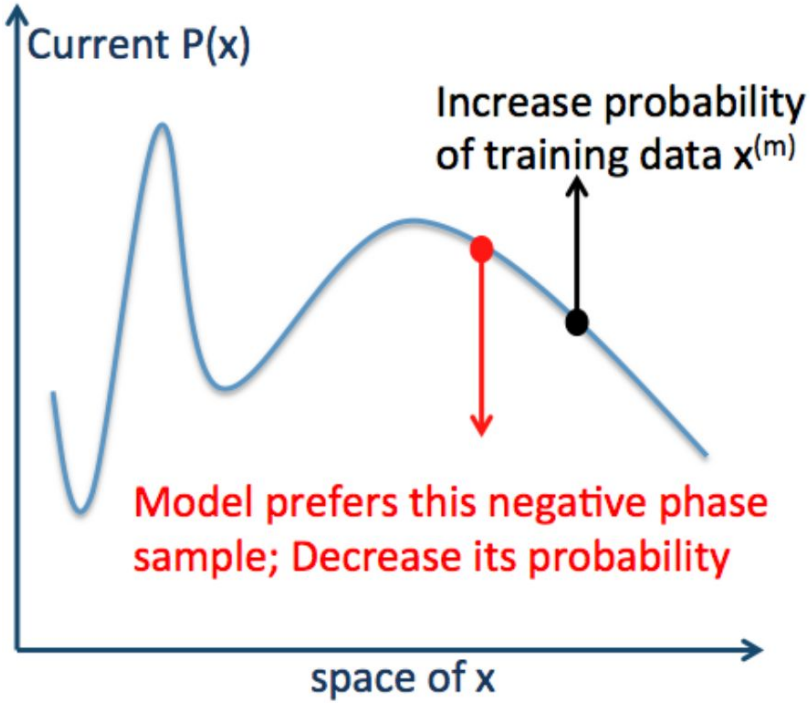
- **The idea:** model the input P(X) better and later focus on optimizing P(Y|X)
- Each layer extract features of the previous
- It removes unimportant variations in the data

RBMs - Pretraining



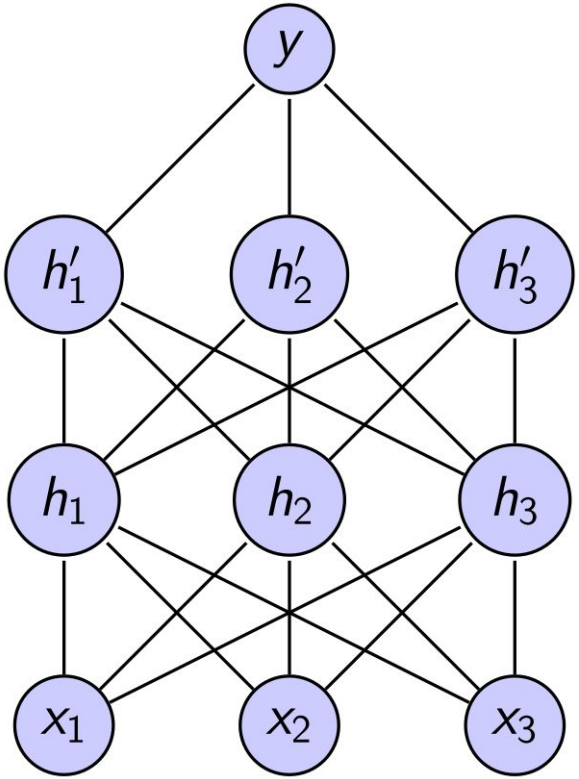
$$\frac{\partial \log p(\mathbf{v})}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}$$

RMBs - Pretraining



$$\frac{\partial \log p(\mathbf{v})}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}$$

RBMs - Pretraining

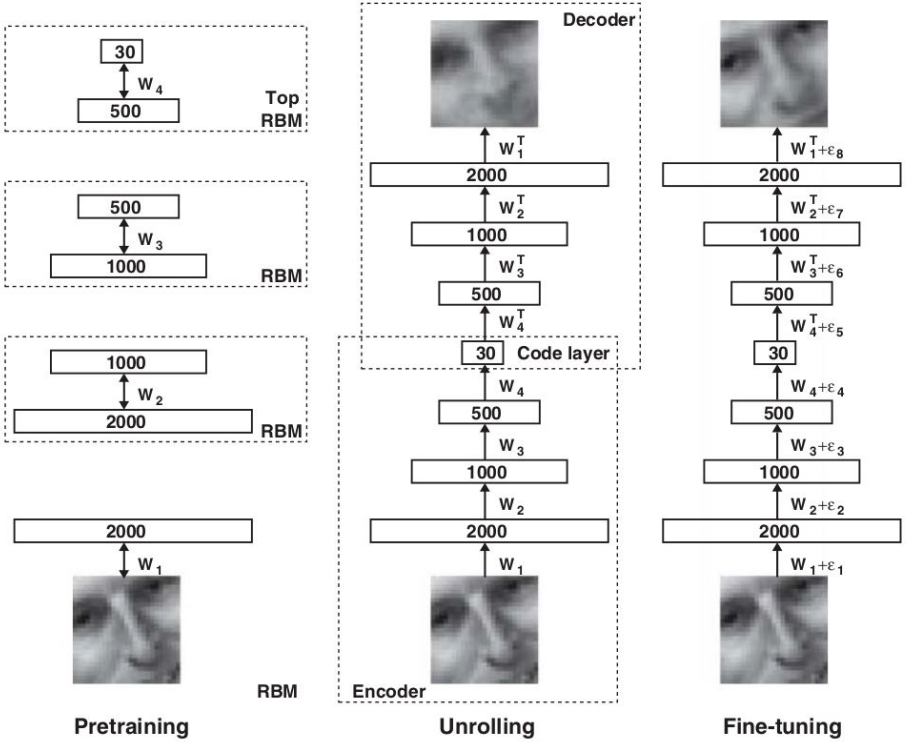


$$\Delta w_{ij} = \epsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model})$$

$$p(h_j = 1 \mid \mathbf{v}) = \sigma(b_j + \sum_i v_i w_{ij})$$

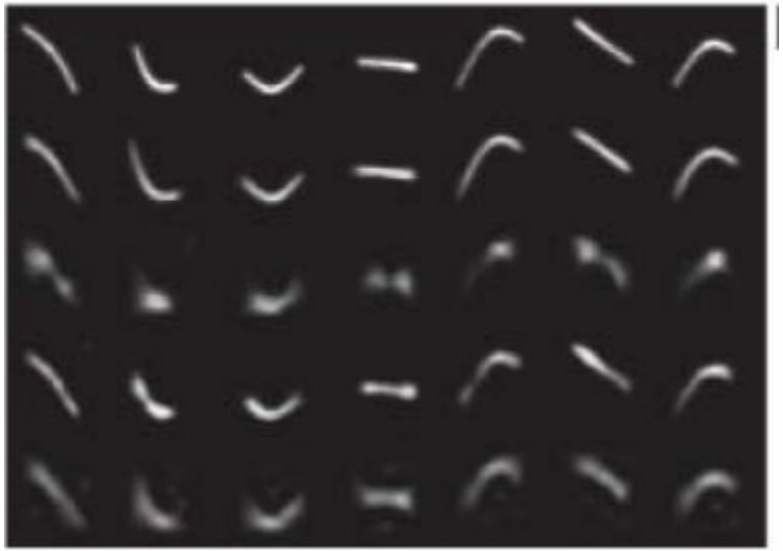
$$p(v_i = 1 \mid \mathbf{h}) = \sigma(a_i + \sum_j h_j w_{ij})$$

Deep Autoencoder



- Train each RBM separately
- “Unfold” the NN to form decoder part
- BP for fine-tuning

Datasets and Results



20000 training and 10000 testing images.

In order: original data, 6-dimensional autoencoder, “logistic” PCA with 6 components, 18 components and standard PCA with 18 components.

Datasets and Results

MNIST dataset

Random test image, 30-dim deep autoencoder, 30-dim logistic PCA, 30-dim PCA

784-1000-500-250-30

Datasets and Results

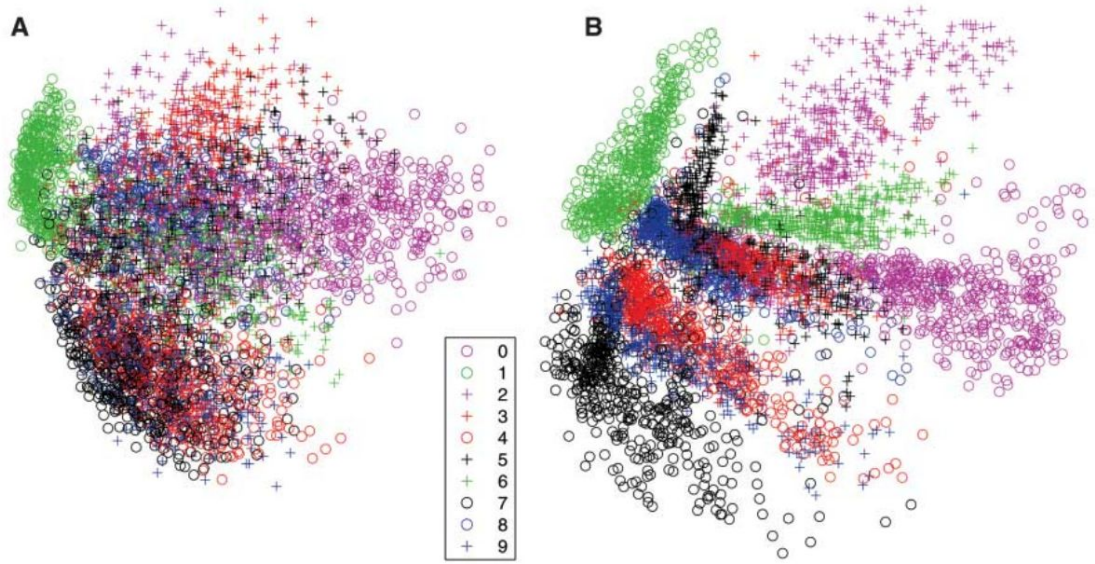
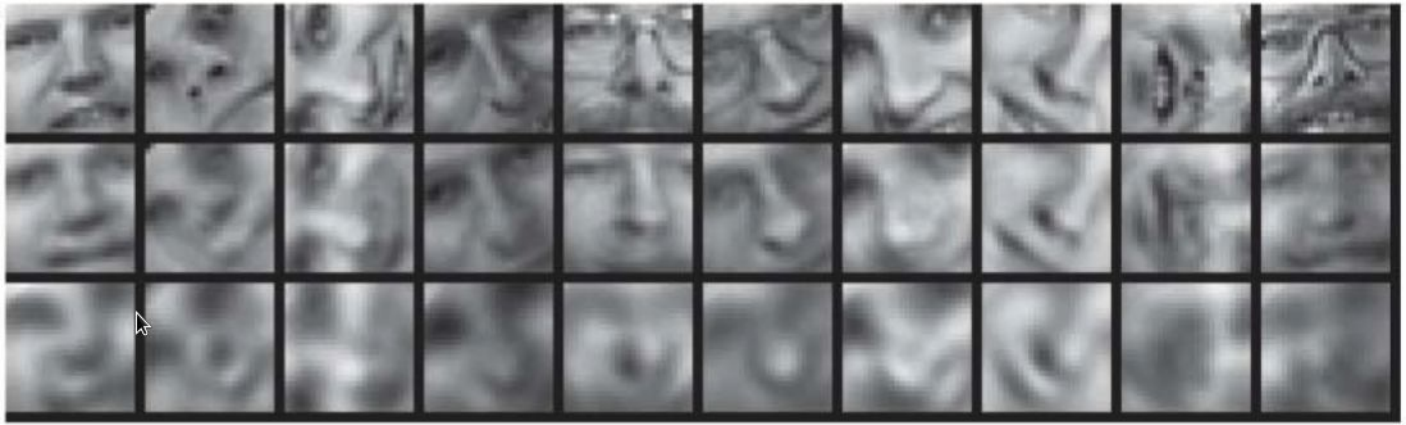


Fig. 3. (**A**) The two-dimensional codes for 500 digits of each class produced by taking the first two principal components of all 60,000 training images. (**B**) The two-dimensional codes found by a 784-1000-500-250-2 autoencoder. For an alternative visualization, see (*8*).

MNIST Dataset

2-dim LSA and autoencoder

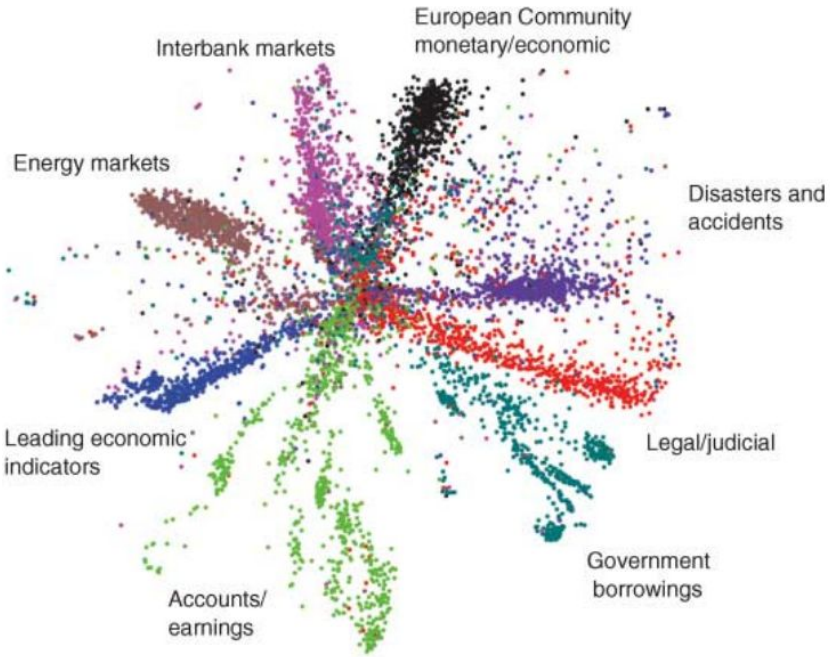
Datasets and Results

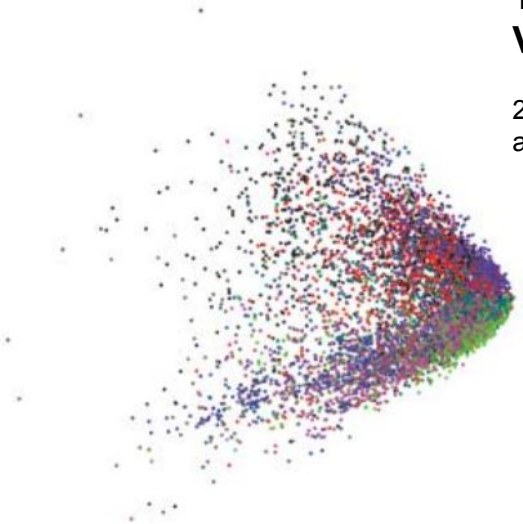


Olivetti Dataset

Random samples of dataset,
625-2000-1000-500-30
autoencoder and 30-dim PCA

Datasets and Results





The Reuters Corpus Volume 2

2000-500-250-125-2 autoencoder and 2-dim LSA

References

G. Hinton, R. R. Salakhutdinov, Science 313, 504-507 (2006)

G. E. Hinton. A practical guide to training restricted Boltzmann machines. Technical Report 2010-003, University of Toronto, 2010.

http://cl.naist.jp/~kevinduh/a/deep2014/