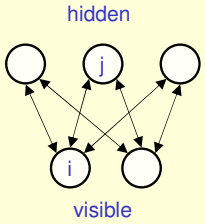
CSC321 Winter 2007

Lecture 21: Some Demonstrations of Restricted Boltzmann Machines

Geoffrey Hinton

A simple learning module: A Restricted Boltzmann Machine

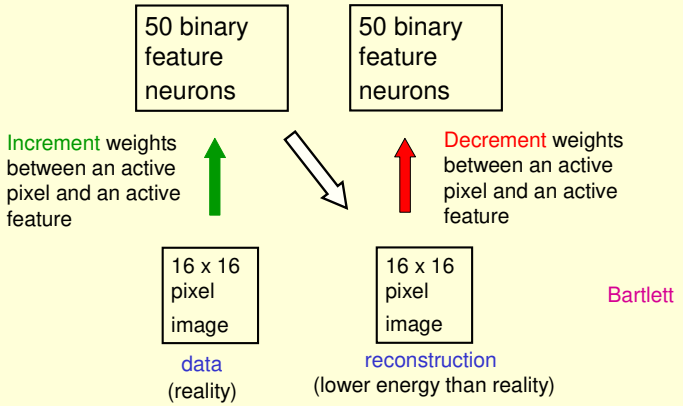
- We restrict the connectivity to make learning easier.
  - Only one layer of hidden units.
    - We will worry about multiple layers later
  - No connections between hidden units.
- In an RBM, the hidden units are conditionally independent given the visible states..
  - So we can quickly get an unbiased sample from the posterior distribution over hidden "causes" when given a data-vector .



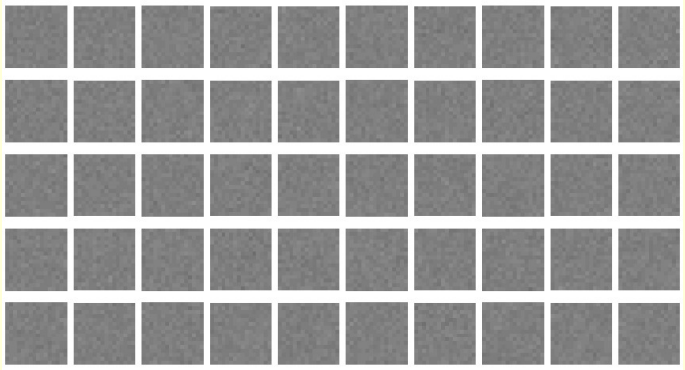
Weights → Energies → Probabilities

- Each possible joint configuration of the visible and hidden units has a Hopfield "energy"
  - The energy is determined by the weights and biases.
- The energy of a joint configuration of the visible and hidden units determines the probability that the network will choose that configuration.
- By manipulating the energies of joint configurations, we can manipulate the probabilities that the model assigns to visible vectors.
  - This gives a very simple and very effective learning algorithm.

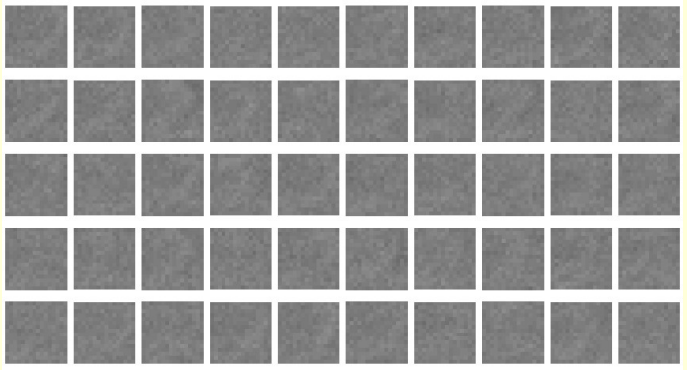
How to learn a set of features that are good for reconstructing images of the digit 2

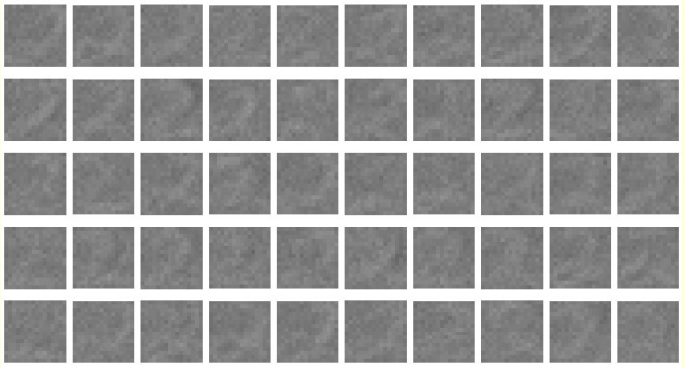


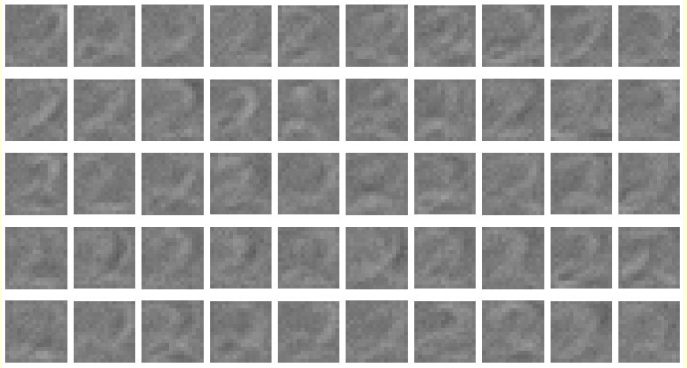
The weights of the 50 feature detectors

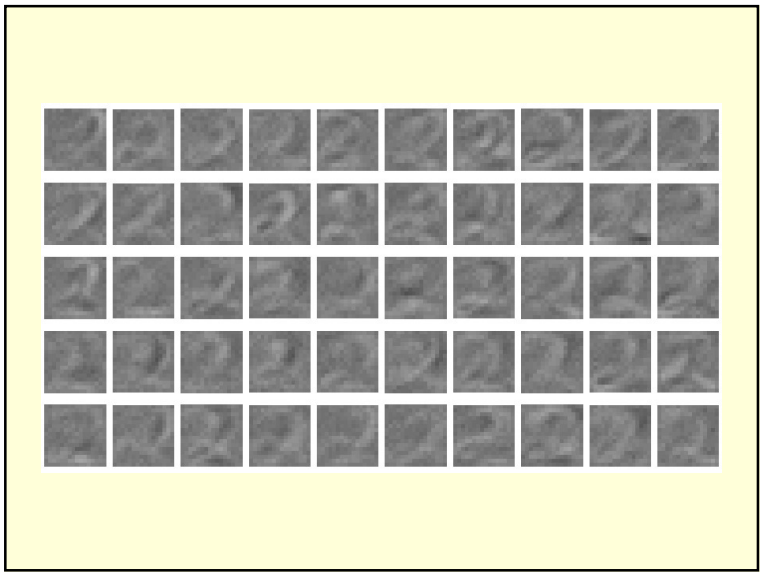


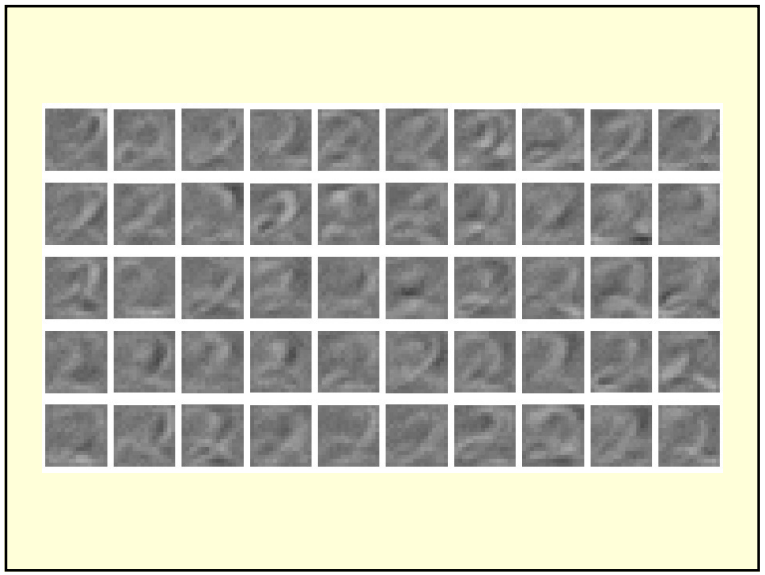
We start with small random weights to break symmetry

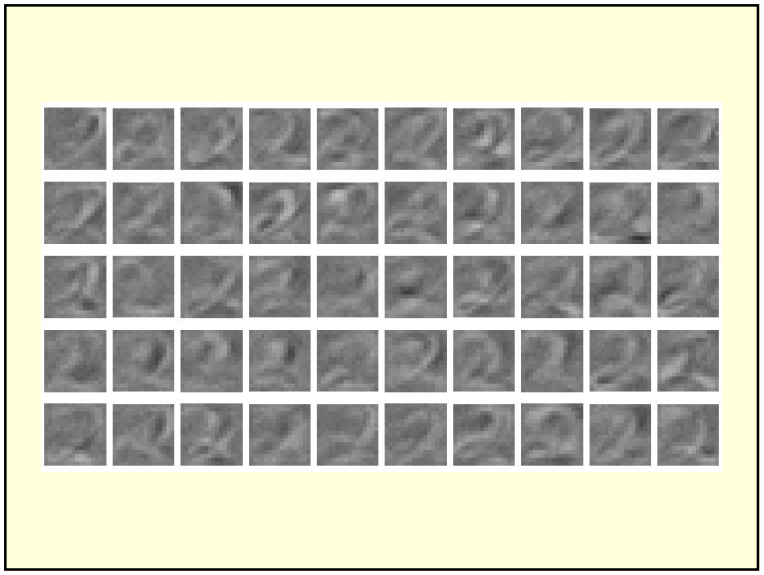


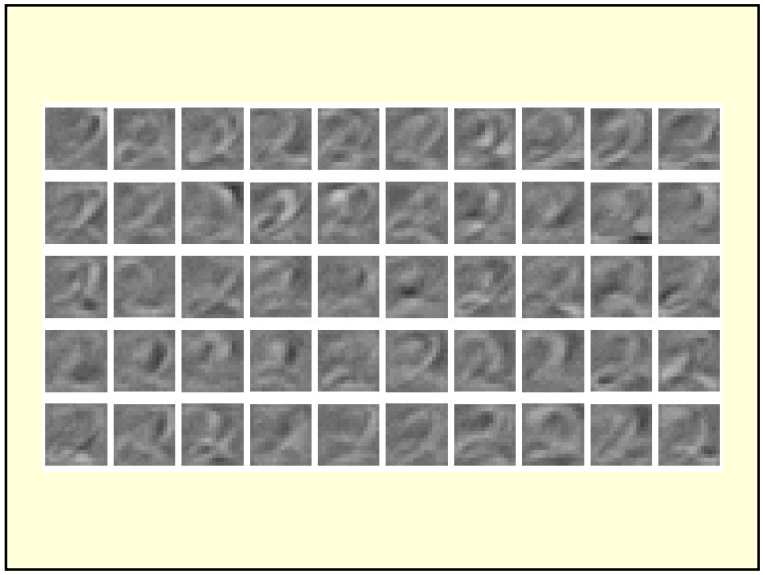


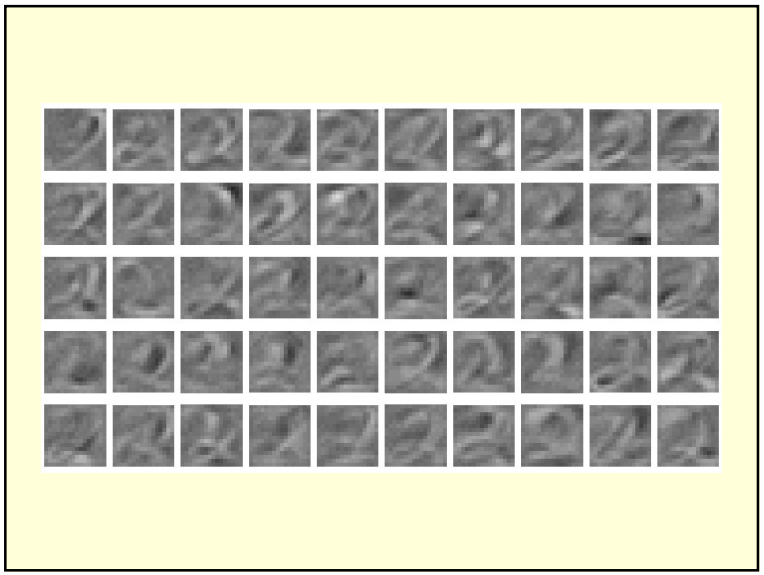


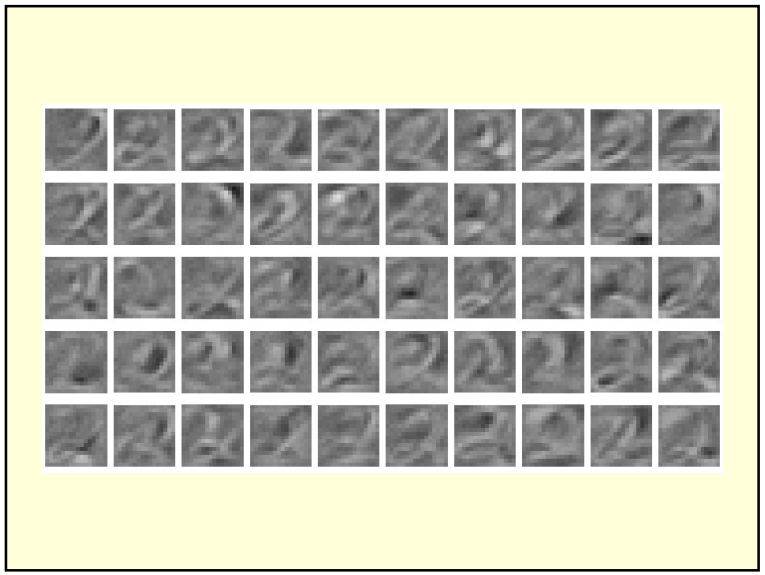


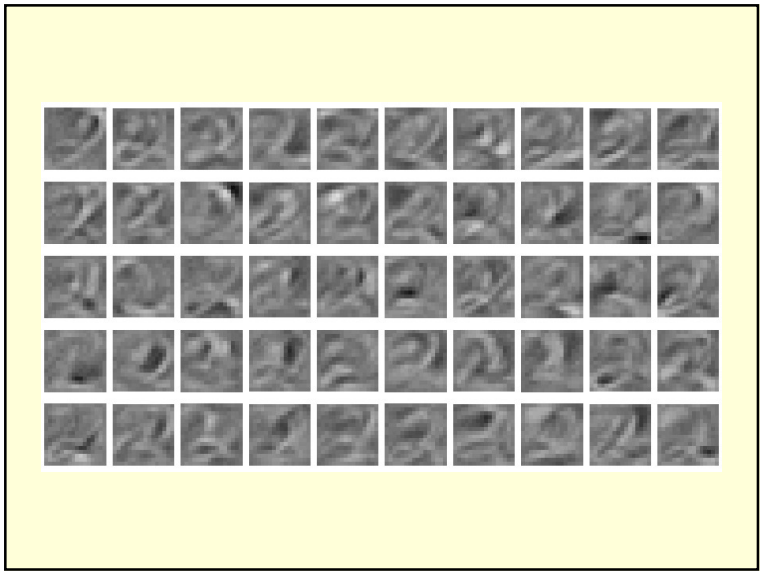


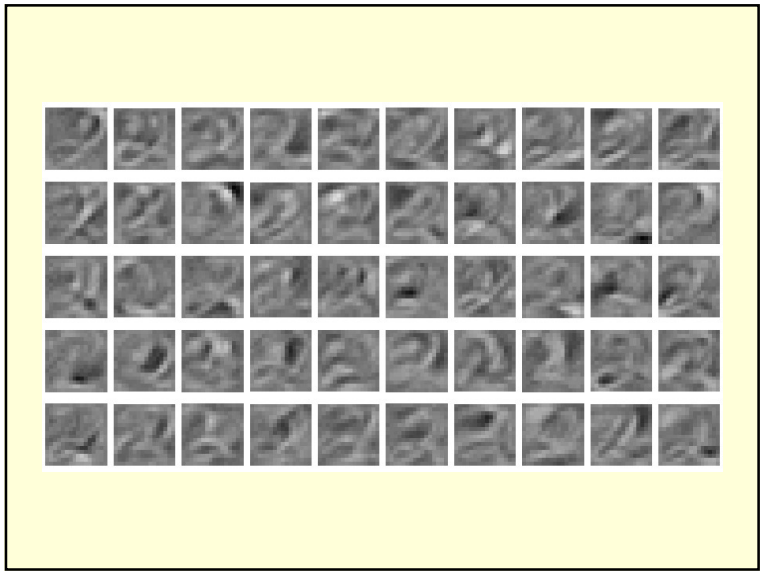


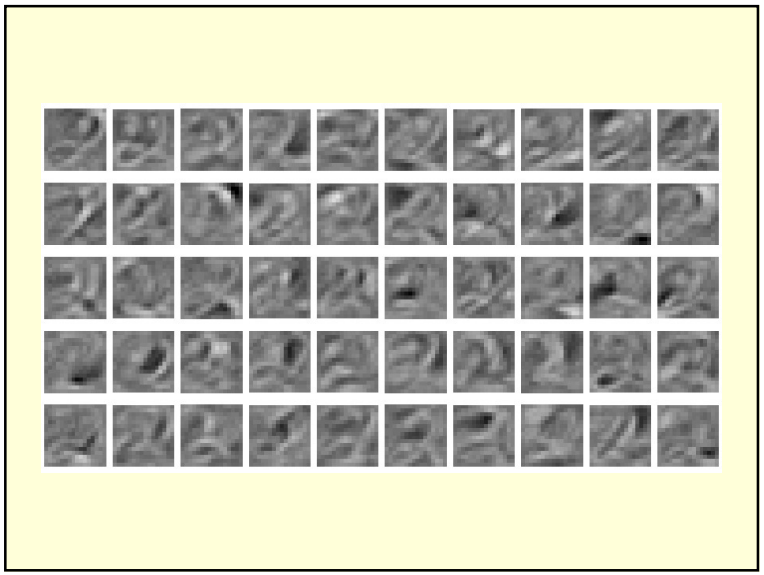


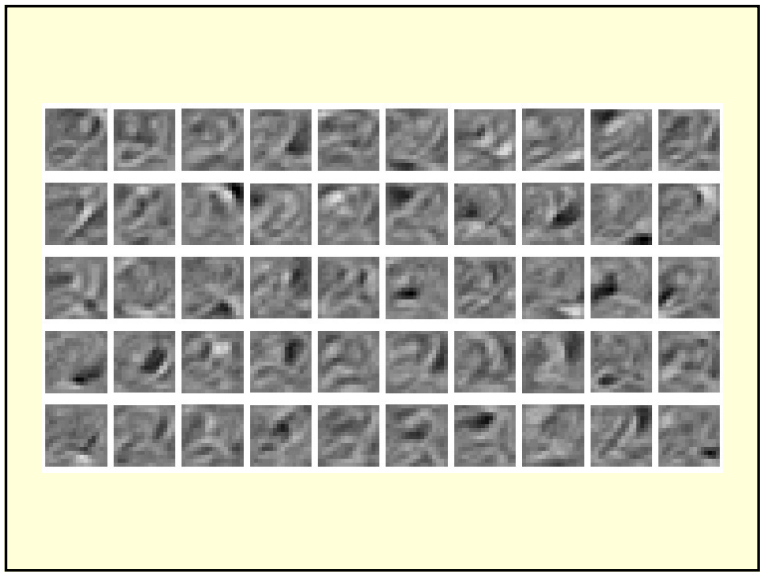


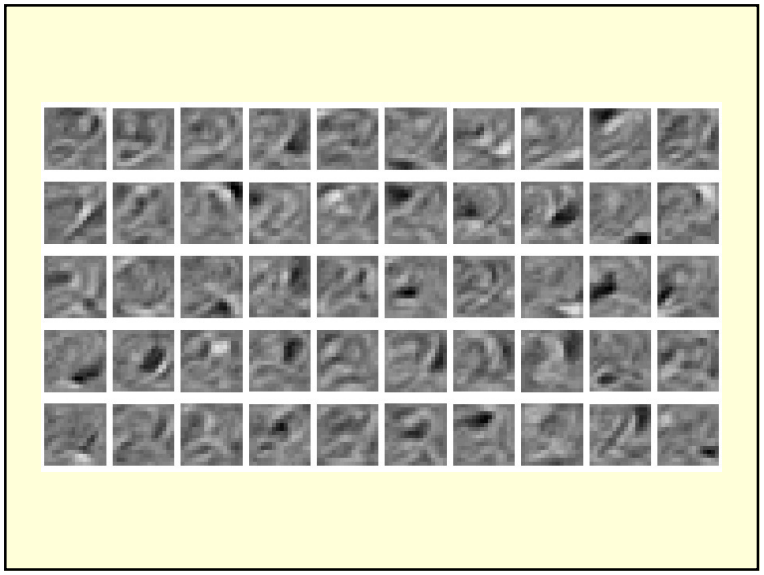


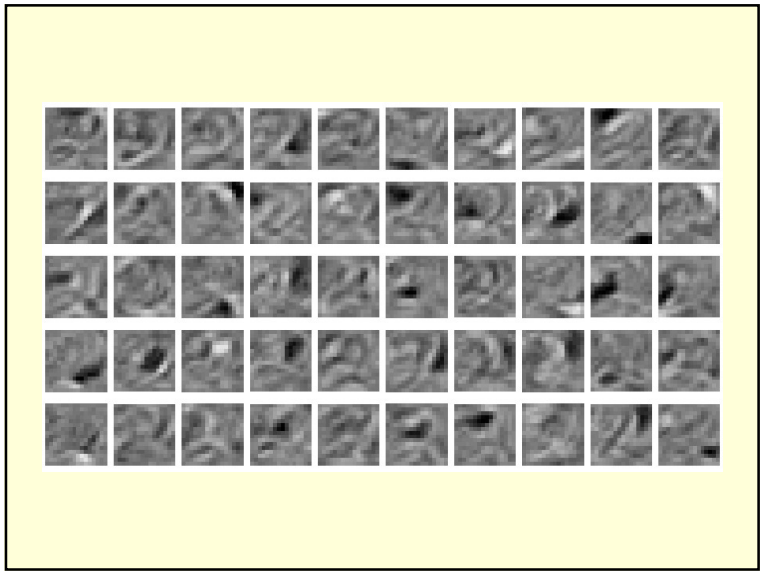


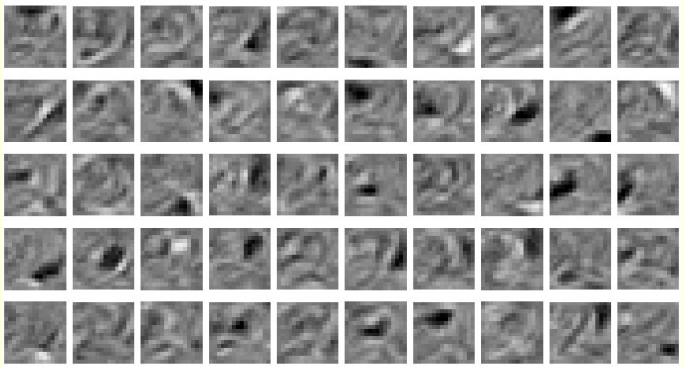




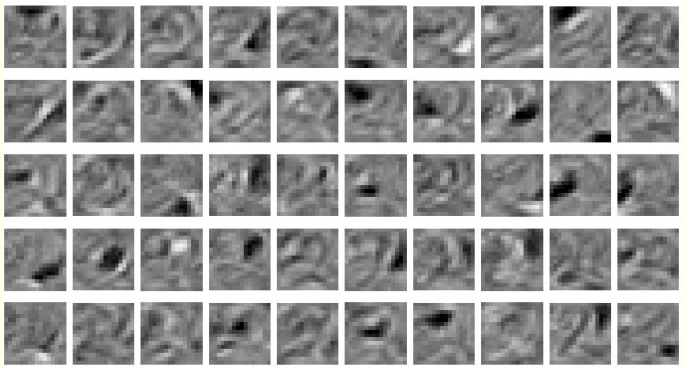




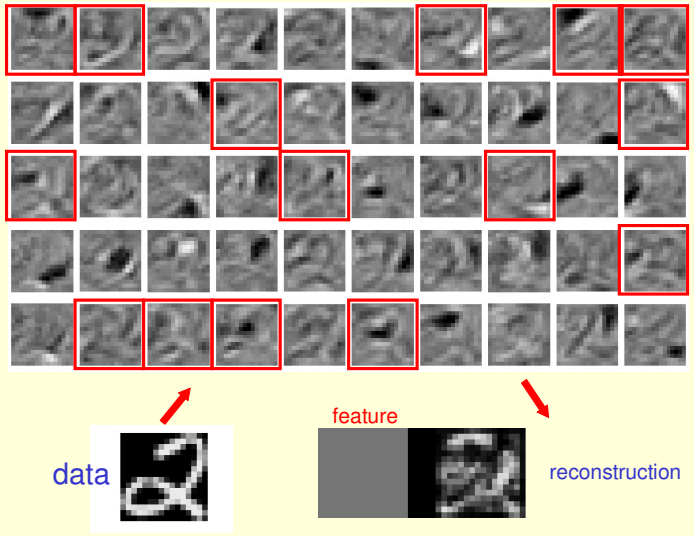




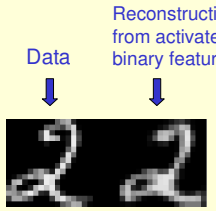
The final 50 x 256 weights



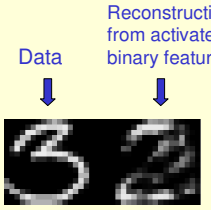
Each neuron grabs a different feature.



How well can we reconstruct the digit images from the binary feature activations?

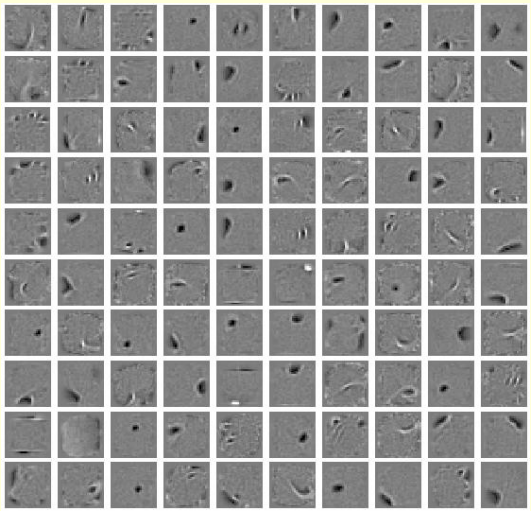


New test images from the digit class that the model was trained on



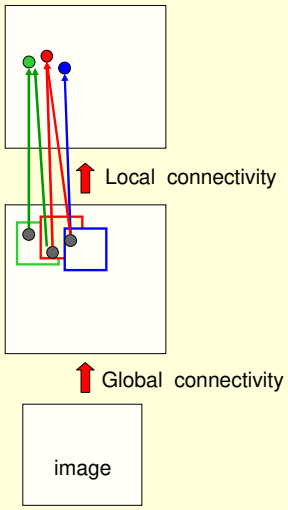
Images from an unfamiliar digit class
(the network tries to see every image as a 2)

Some features learned in the first hidden layer for all digits



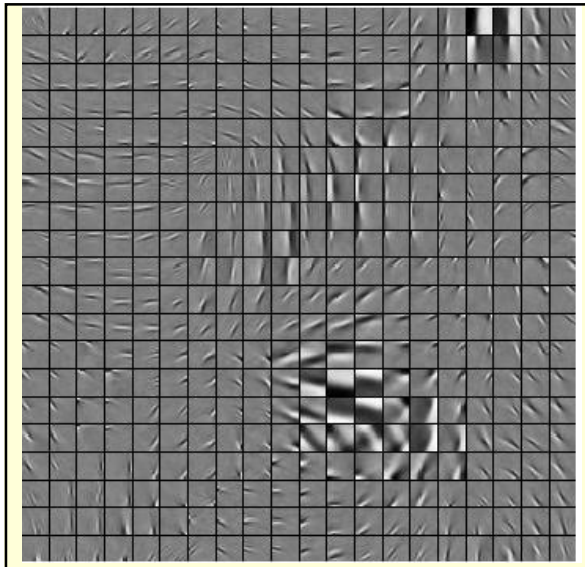
And now for something a bit more realistic

- Handwritten digits are convenient for research into shape recognition, but natural images of outdoor scenes are much more complicated.
  - If we train a network on patches from natural images, does it produce sets of features that look like the ones found in real brains?



A network with local connectivity

The local connectivity between the two hidden layers induces a topography on the hidden units.



Features learned by a net that sees 100,000 patches of natural images.

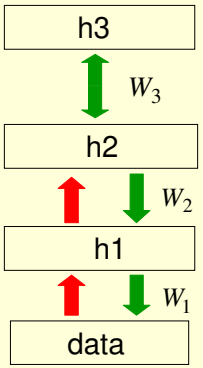
The feature neurons are locally connected to each other.

Osindero, Welling and Hinton (2006) Neural Computation

Training a deep network

- First train a layer of features that receive input directly from the pixels.
- Then treat the activations of the trained features as if they were pixels and learn features of features in a second hidden layer.
- It can be proved that each time we add another layer of features we get a better model of the set of training images.
  - i.e. we assign lower energy to the real data and higher energy to all other possible images.
  - The proof is complicated. It uses variational free energy, a method that physicists use for analyzing complicated non-equilibrium systems.
  - But it is based on a neat equivalence.

The generative model after learning 3 layers

- To generate data:
  1. Get an equilibrium sample from the top-level RBM by performing alternating Gibbs sampling.
  2. Perform a top-down pass to get states for all the other layers.

So the lower level bottom-up connections are not part of the generative model