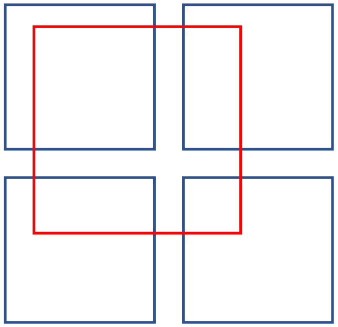
CNN Examples - Different Complexities

- LeNet
- AlexNet

LeNet

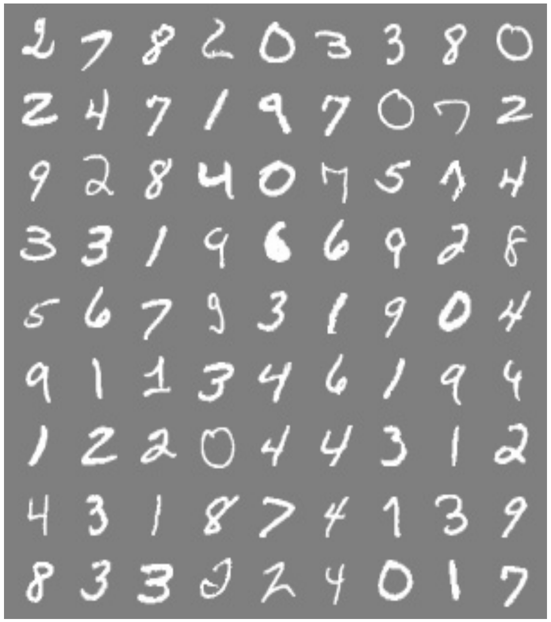
Layer Number	Input Shape	Receptive Field	Number of Feature Maps	Type of Neuron
1	28 X 28 X 1	5 X 5	20	Convolutional
2	24 X 24 X 20	2 X 2		Pooling
3	12 X 12 X 20	5 X 5	50	Convolutional
4	8 X 8 X 50	2 X 2		Pooling
5	800	1 X 1	500	Fully Connected
6	500		10	Softmax

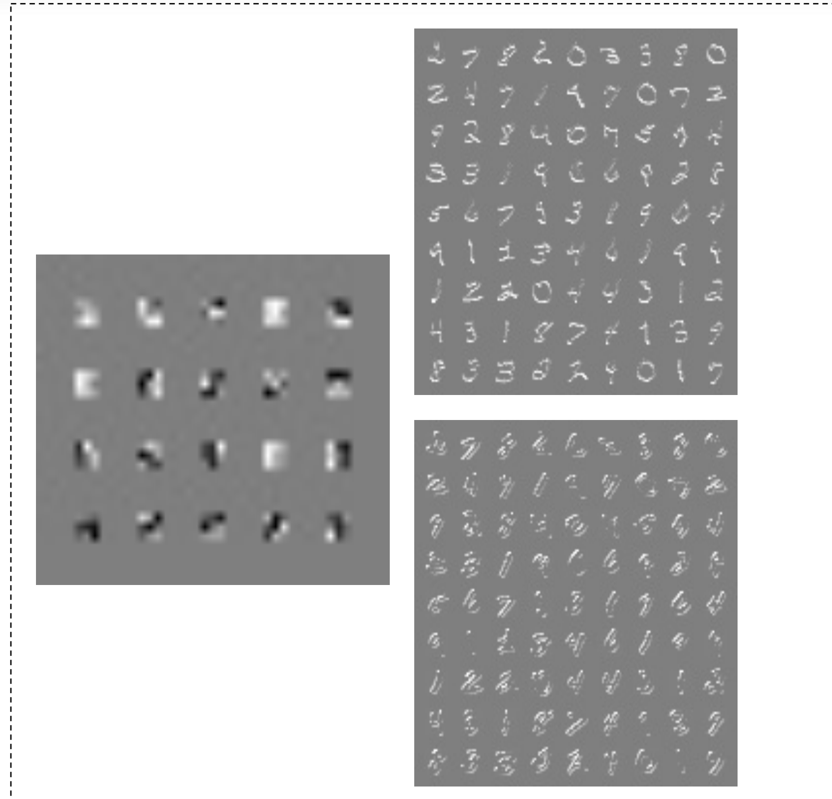
- Each pixel in layer 3 corresponds to 7/3 of a pixel in the input Second level
- Receptive field of layer 1 is 5X5



LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278-2324.

Case Study: LeNet

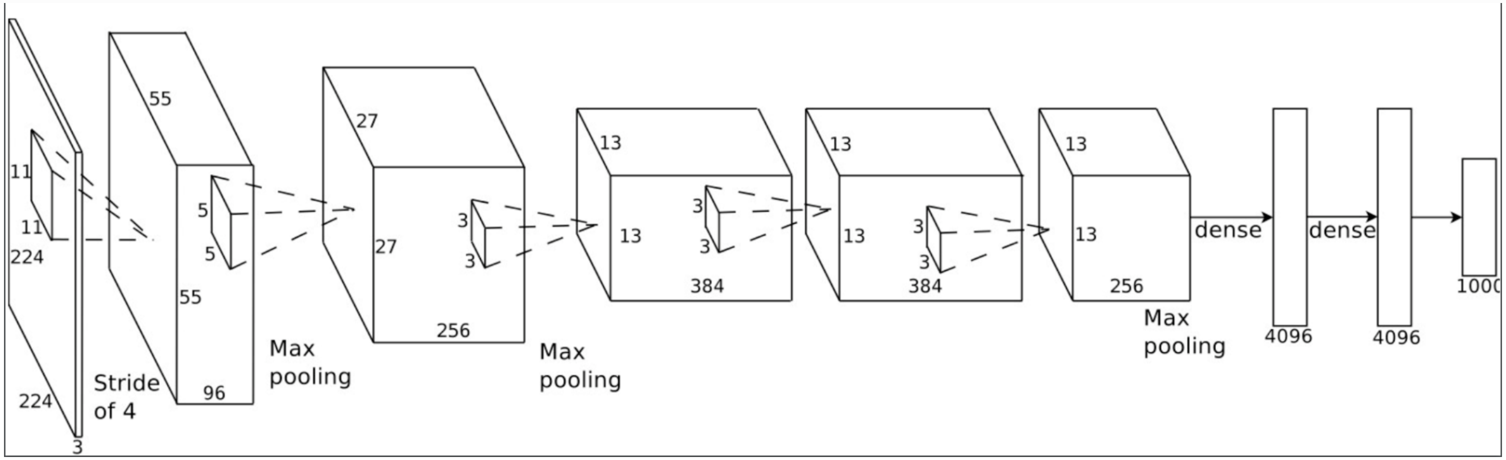


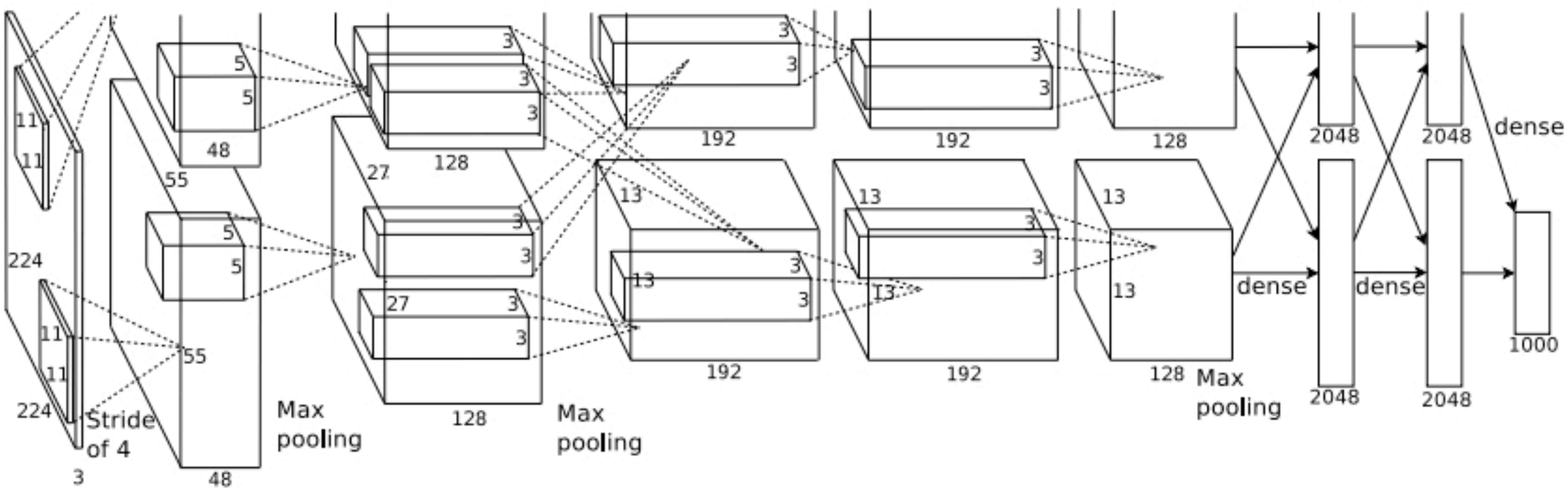


- This is after training the network for 75 epochs with a learning rate of 0.01

- Produces an accuracy of 99.38% on the MNIST dataset.

Case Study: AlexNet





Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

Case Study: AlexNet - 1 of 3

Layer Number	Input Shape	Receptive Field	Number of Kernels	Type of Neuron
1	224 X 224 X 3	11 X 11, stride 4	96	Convolutional
2		3 X 3, stride 2		Pooling
3	55 X 55 X 96	5 X 5	256	Convolutional
4		3 X 3, stride 2		Pooling
5	13 X 13 X 256	3 X 3, padded	384	Convolutional
6	13 X 13 X 384	3 X 3, padded	384	Convolutional
7	13 X 13 X 384	3 X 3	256	Convolutional
8	43264	1 X 1	4096	Fully Connected
9	4096	1 X 1	4096	Fully Connected
10	4096		1000	Softmax

Receptive field of the layer 7 is

~ 52 pixels !! which is almost as big as an object part (about one – fourth of the input image)

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

Case Study: AlexNet - 2 of 3

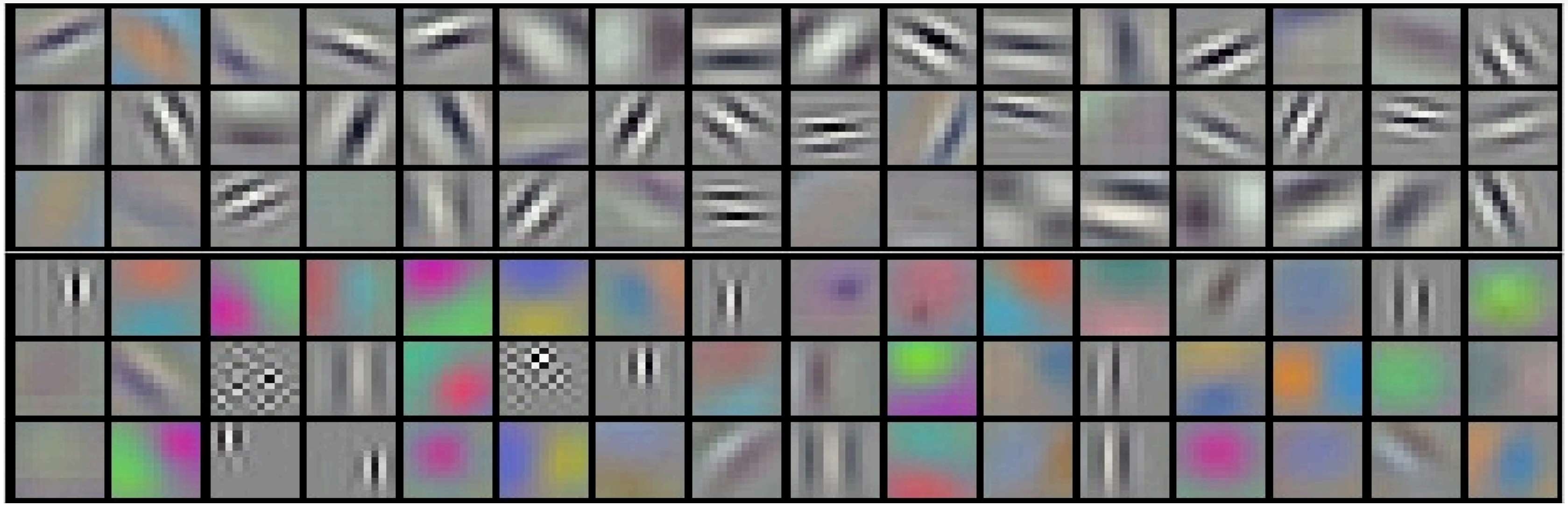
- Imagenet -15 million images in over 22,000 categories
- Imagenet categories are much more complicated than other datasets
  - Often difficult even for humans to categorize perfectly
  - Average human-level performance is about 96% on this dataset
- (ILSVRC), used about 1000 of these categories
- AlexNet was the earliest systems to break the 80% mark
  - Non-neural conventional techniques were unable to achieve such performance

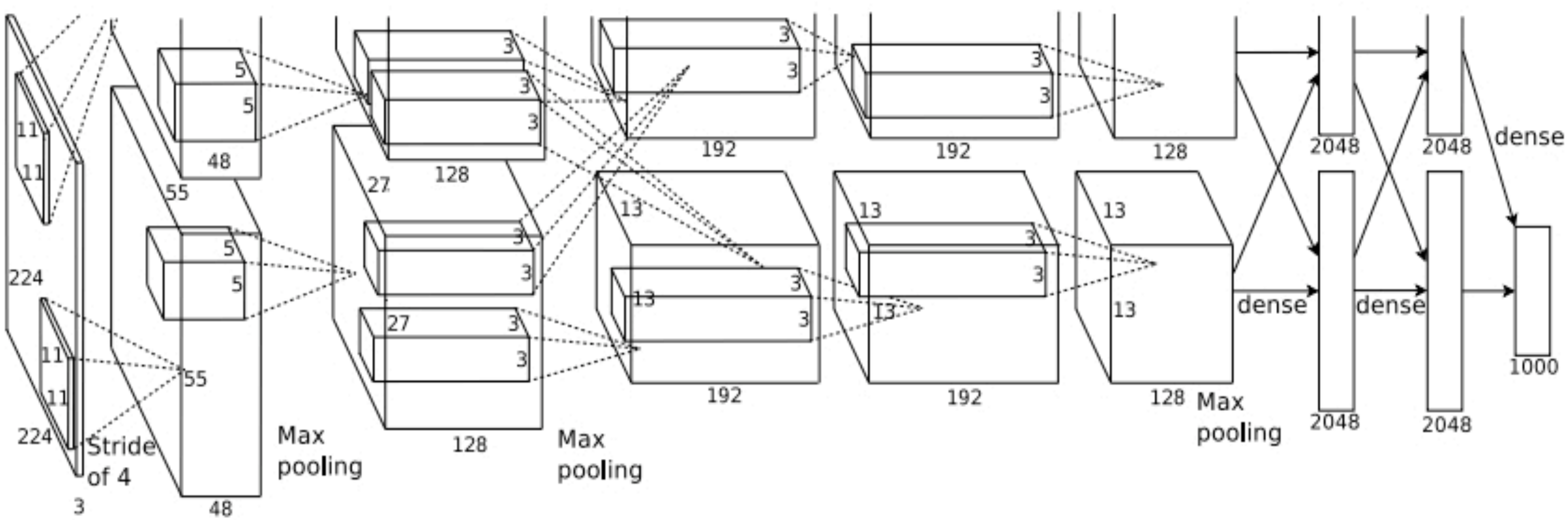
Case Study: AlexNet - 3 of 3

- AlexNet was huge at the time.
  - The size could lead to instability during training or inability to learn, if without proper regularization

- Some techniques were used to make it trainable
  - AlexNet was the first prominent network to feature ReLU
  - Features multi-GPU training (originally trained the networks on two Nvidia GTX 580 GPUs with 3GB)

Case Study: AlexNet Filters





Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

CNN Recap

The CNNs are similar to the basic MLP architecture illustrated earlier, but some key extensions include:

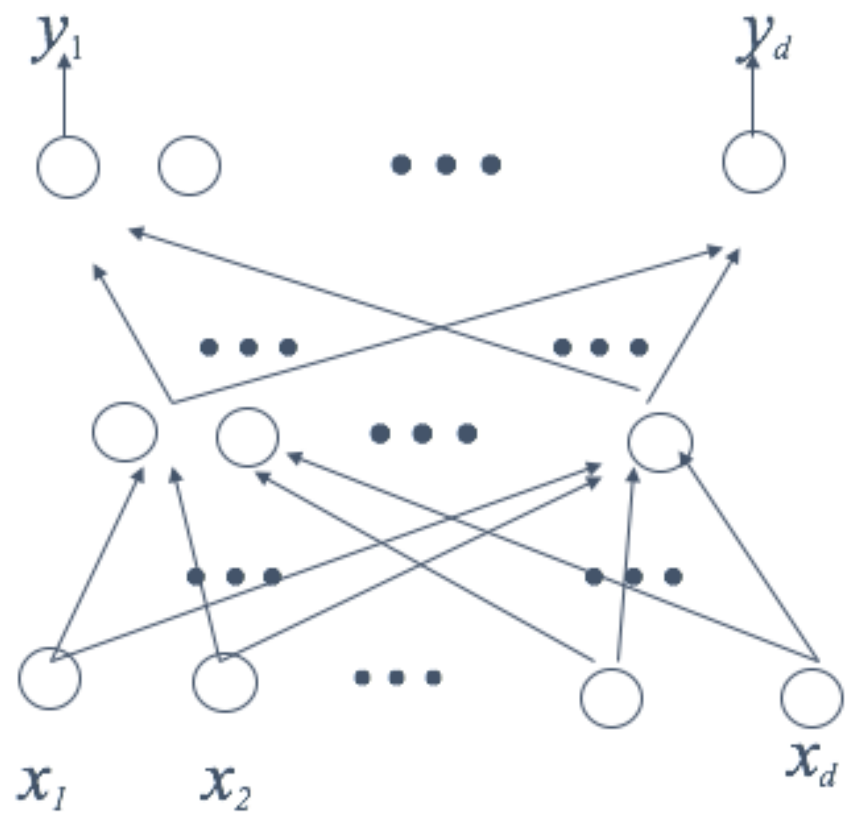
- The concept of weight-sharing through kernels
- Weight-sharing enables learnable kernels, which in turn define feature maps
- The idea of pooling

Auto-encoder - 1 of 4

- Networks seen thus far are all trained via supervised learning

- Sometimes we may need to train a network without supervision:
  - → Unsupervised learning

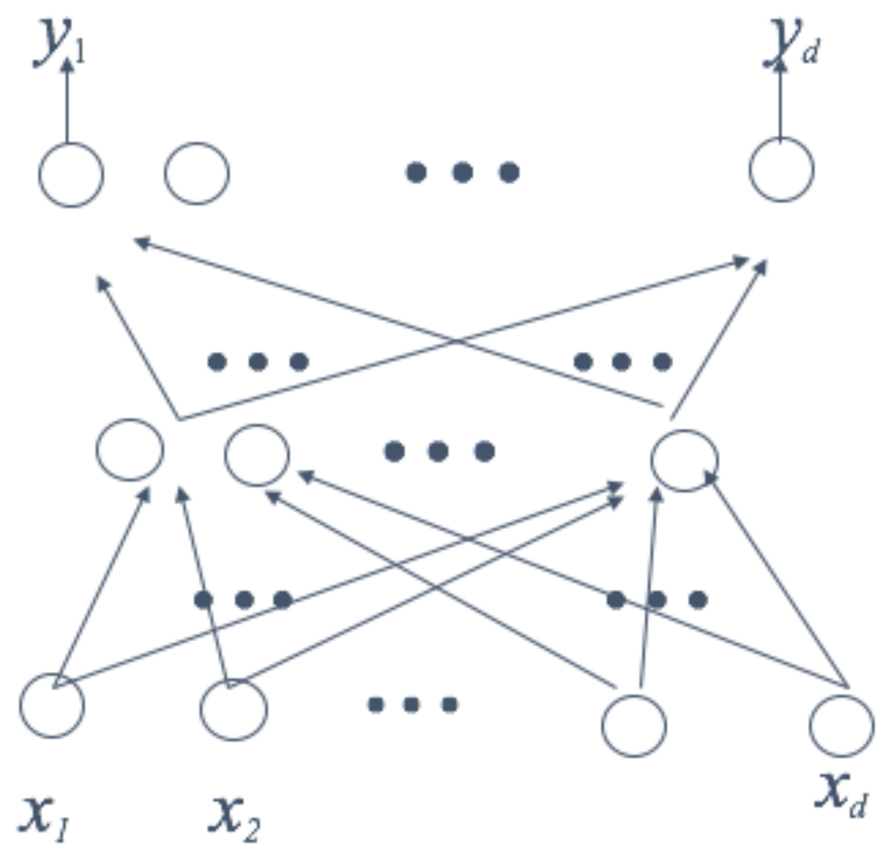
- Auto-encoder is a such example
  - Consider y_i being an approximation of x_i.



Auto-encoder - 2 of 4

- Perfect auto-encoder would map x_i to x_i

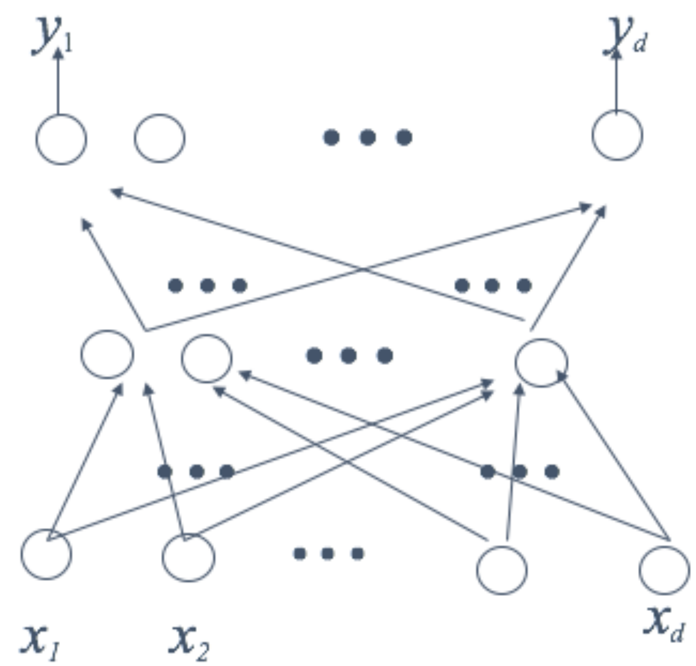
- Learn good representations in the hidden layer



Auto-encoder - 3 of 4

- Consider two cases
  1. Much fewer hidden nodes than input nodes
  2. Many hidden nodes or more hidden nodes than input nodes

- **Case 1:** Encoder for compressing input and compressed data should still be able to reconstruct the input
  – Similar to, e.g., PCA



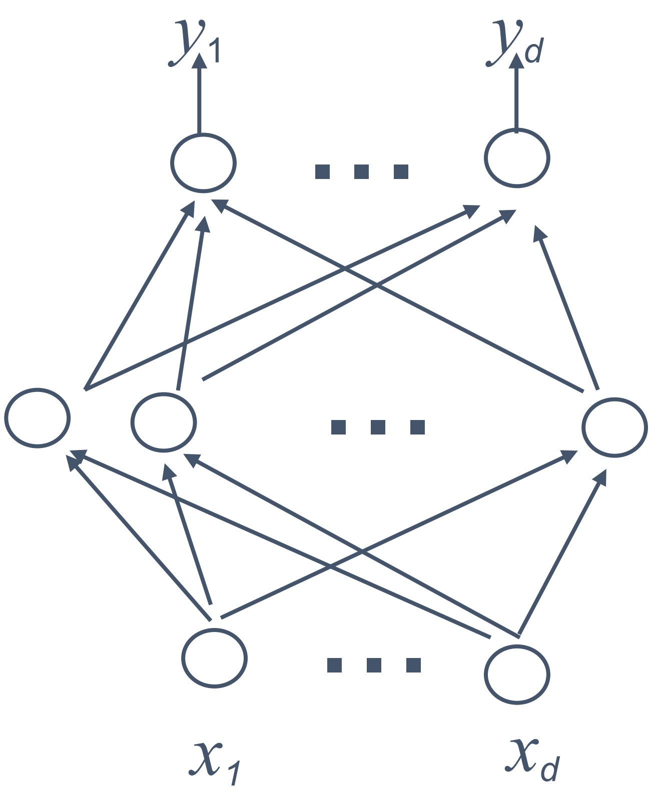
Auto-encoder - 4 of 4

Consider two cases

1. Fewer hidden nodes than input nodes
2. More hidden nodes than input nodes

Case 2: Allow more hidden nodes than input

- Allow more freedom for the input-to-hidden layer mapping in exploring structure of the input
- Additional “regularization” will be needed in order to find meaningful results



Recurrent Neural Networks (RNNs) – 1/4

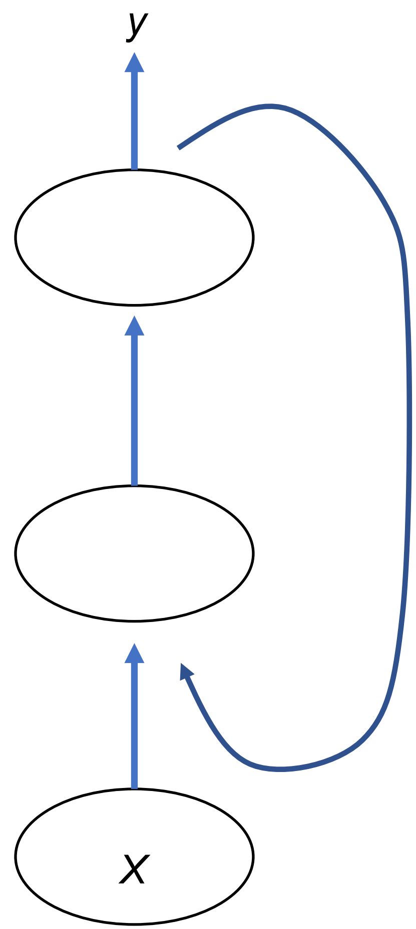
- *Feedforward networks*: Neurons are interconnected without any cycle in the connection

- *Recurrent neural networks*: Allow directed cycles in connections between neurons
  - Notion of “state” or temporal dynamics
  - Necessity of internal memory

- One clear benefit: Such networks could naturally model variable-length sequential data

RNNs – 2 of 4

- A basic, illustrative architecture for RNN (showing only one node each layer)
  - **QUESTION:** What is this network equivalent to, if we "unfold" the cycles for a given sequence of data?



RNNs – 3 of 4

- Training with BP algorithm may suffer from so-called *vanishing gradient problem*
- Some RNN variants have sophisticated “recurrence” structures, invented in part to address such difficulties faced by basic RNN models

RNNs – 4 of 4

Examples:

- The "Long short-term memory" (LSTM) model
  - used to produce state-of-the-art results in speech and language applications
- The Gated Recurrent Unit model, illustrated here: