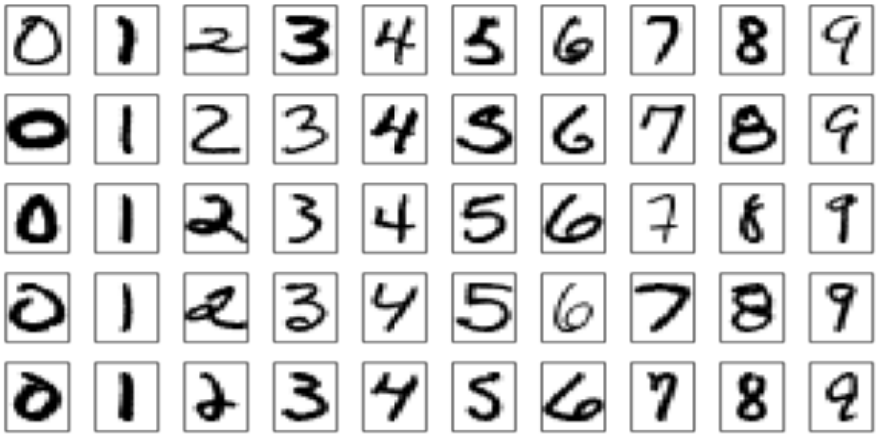
Modern Machine Learning — Neural Networks Example

Kenneth E. Barner
Department of Electrical and Computer Engineering
University of Delaware

Example: Zip Code Data



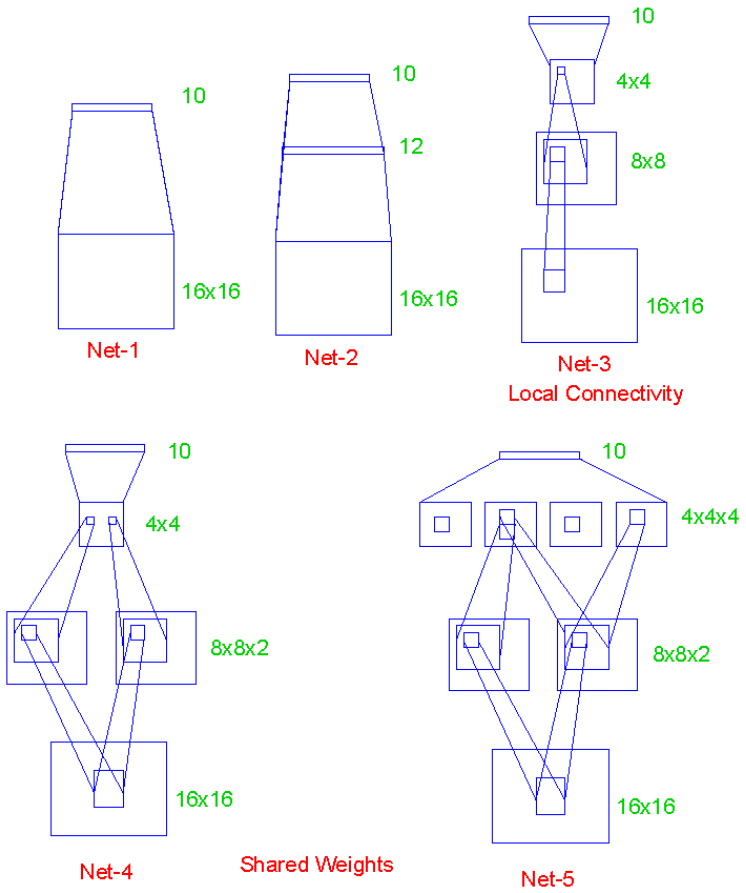
Example of training cases from ZIP code data. Each image is a 16x16 8-bit grayscale representation of a handwritten digit.

Data: Normalized handwritten digits, automatically scanned from envelopes by the U.S. Postal Service

Images are 16 x 16, with 320 digits in the training set and 160 in the test set

Goal: Recognize digits in the test set

Example: Zip Code Data



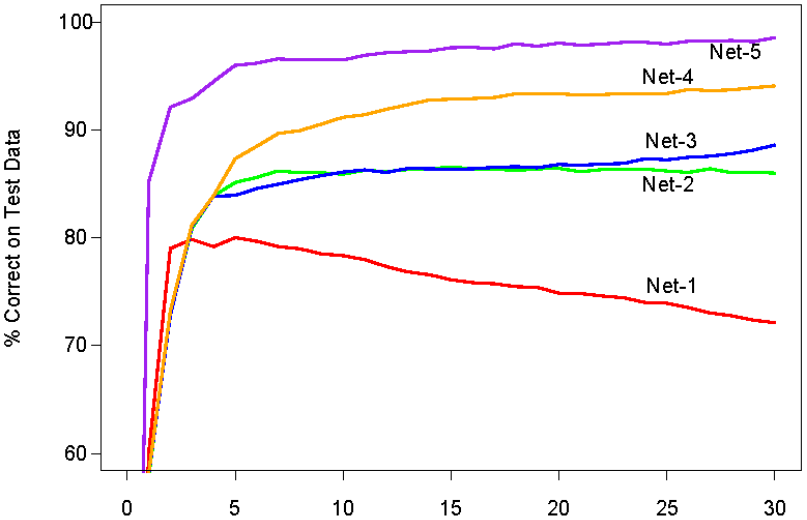
Architecture of the five networks used in the ZIP code example.

Five different architecture networks were implemented:

- Net-1: No hidden layer, equivalent to multinomial logistic regression
- Net-2: One hidden layer, 12 hidden units fully connected
- Net-3: Two hidden layers locally connected
- Net-4: Two hidden layers, locally connected with weight sharing
- Net-5: Two hidden layers, locally connected, two levels of weight sharing

Note: All implementations utilized sigmoidal activation functions and the sum-of-squares cost function for optimization

Example: Zip Code Data



Test performance curves, as a function of the number of training epochs

Test set performance of five different neural networks on a handwritten digit classification example.

Network Architecture	Links	Weights	% Correct
Net-1: Single layer network	2570	2570	80.0%
Net-2: Two layer network	3214	3214	87.0%
Net-3: Locally connected	1226	1226	88.5%
Net-4: Constrained network 1	2266	1132	94.0%
Net-5: Constrained network 2	5194	1060	98.4%

Observations:

- The number of parameters is > the number of training samples ⟹ error on the training text is 0%
- Net-1 has no hidden layer, is nearly equivalent to multinomial regression; training over fits the data
- Net-2 is a traditional single hidden layer network
- Net-3, 4, & 5 introduced structures and constraints tailored to the problem at hand — reducing parameters while improving performance
- Net-3 uses patches and local connectivity to extract local features
- Net-4 & 5 utilize local connectivity and shared weights; local units perform the same function on different parts of the image (because of shared weights)
- Shared weights yield extracted features in different parts of the image to be computed by the same linear function; such networks are known as convolutional networks