UNIVERSITAT POLITÈCNICA
DE CATALUNYA

FIB

Some applications of MLPs trained with backpropagation

MACHINE LEARNING/
APRENENTATGE (A)

Lluís A. Belanche

Year 2010/11

Sonar target recognition

(Gorman and Sejnowski, 1988)

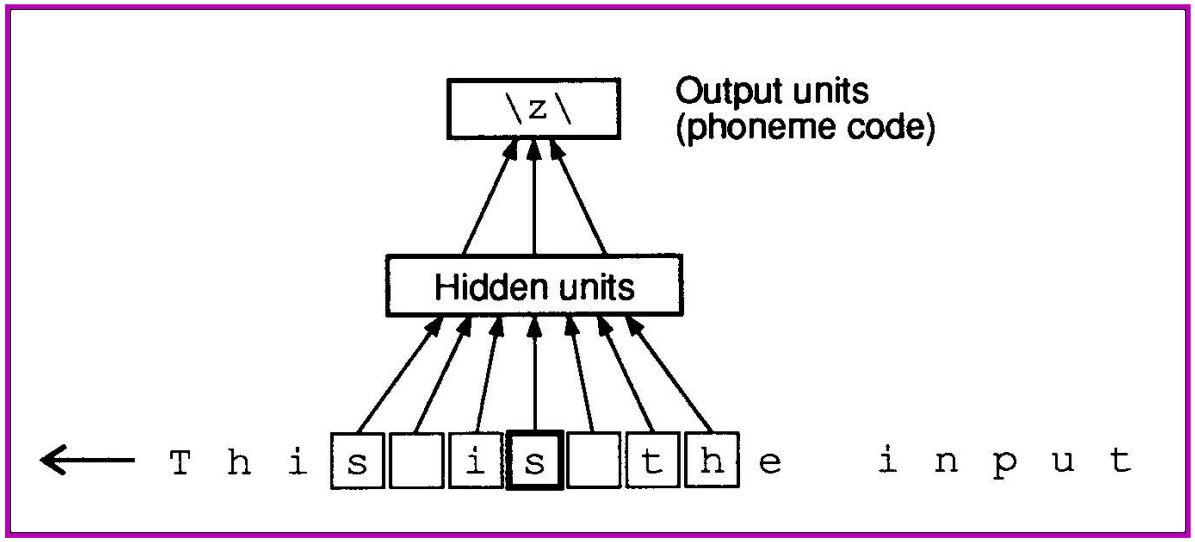
- Two-layer backprop network trained to distinguish between reflected sonar signals of rocks and metal cylinders at bottom of Chesapeake Bay
- 60 input units, 2 output units
- Input patterns based on Fourier transform of raw time signal
- Tried varying numbers of hidden units {0, 3, 12, 24}
- Best performance is obtained with 12 hidden units (close to 100% training set accuracy)
- Test set accuracy is 85-90%

NETTalk

(Sejnowski & Rosenberg, 1987 "*Parallel Networks that Learn to Pronounce English Text*", Complex Systems 1, 145-168)

- Project for pronouncing English text: for each character, the network should give the code of the corresponding phoneme:
  - A stream of words is given to the network, along with the phoneme pronunciation of each in symbolic form
  - A speech generation device is used to convert the phonemes to sound
- The same character is pronounced differently in different contexts:
  - H<u>e</u>ad
  - B<u>e</u>ach
  - L<u>e</u>ech
  - Sk<u>e</u>tch

NETTalk – the architecture



- Input is rolling sequence of 7 characters
- 7 x 29 possible characters = 203 binary inputs
- 80 neurons in one hidden layer
- 26 output neurons (one for each phoneme code)
- 16,240 weights in the first layer; 2,080 in the second

➔ 203-80-26 two-layer network

NETTalk – Training results

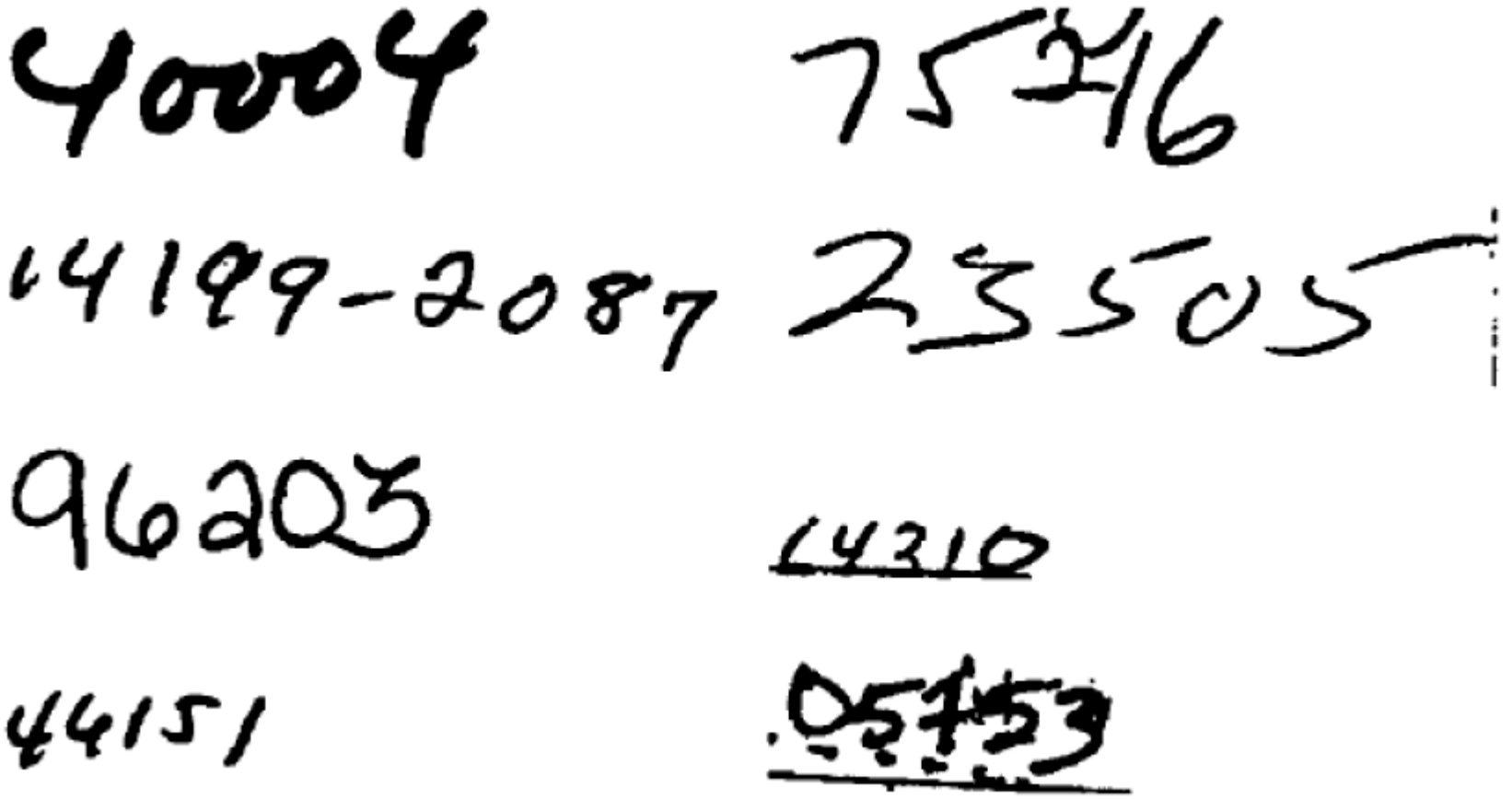
- Training set: database of 1,024 words
- After 10 epochs the network obtains intelligible speech; after 50 epochs 95% accuracy is achieved
  - generalization: 78% accuracy on continuation of training text
  - Since three characters on each side are not always enough to determine the correct pronunciation, 100% accuracy cannot be obtained
- The learning process
  - Gradually performs better and better discrimination
  - Sounds like a child learning to talk
  - damaging network produced graceful degradation, with rapid recovery on retraining
- Analysis of the hidden neurons reveals that some of them represent meaningful properties of the input (e.g., vowels vs. consonants)

NETTalk
Comparison to Rule-Based

- Generalization of NETTalk: only 78% accuracy
- Tools based on hand-coded linguistic rules (e.g., DECtalk) achieve much higher accuracy
- Hand-coded linguistic rules developed over a decade, and were worth thousands of $
- “Flagship” demonstration that converted many scientists, particularly psychologists, to neural network research
- The data for NETTalk used to be found at:

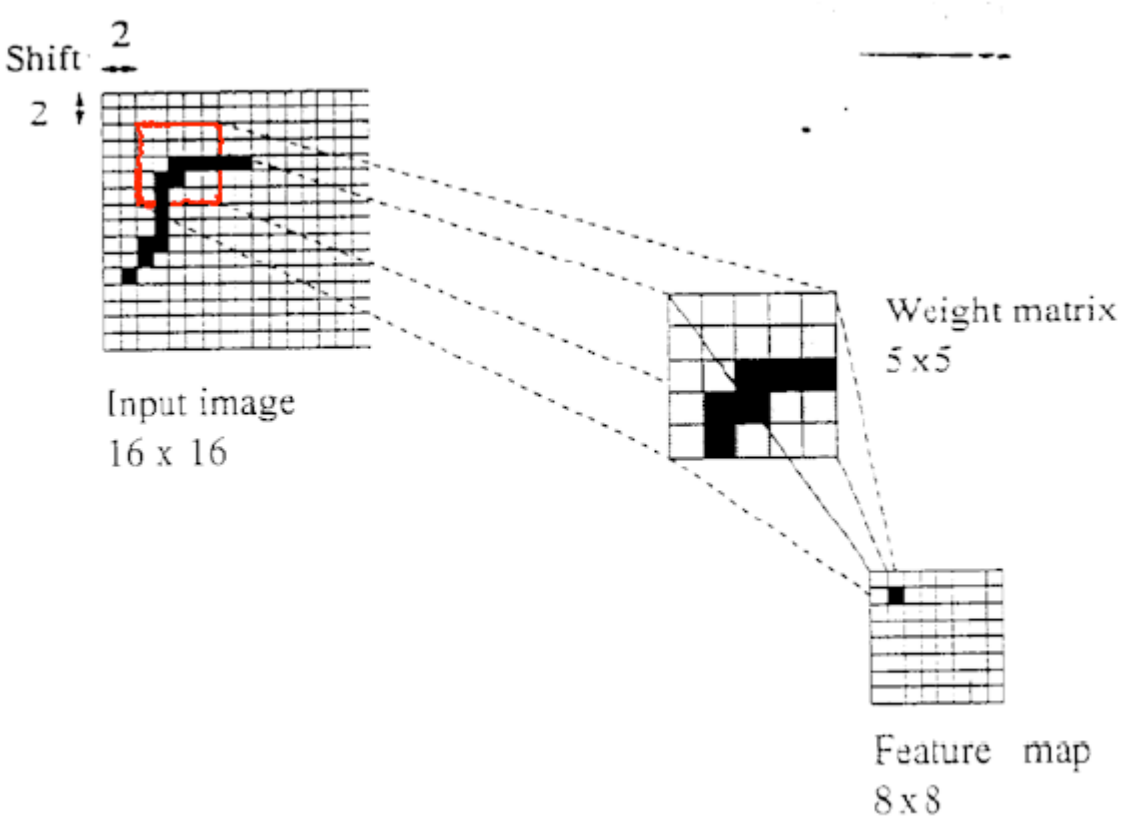
  http://homepages.cae.wisc.edu/~ece539/data/nettalk/

Zipcode Recognition (Y. LeCun, 1990)

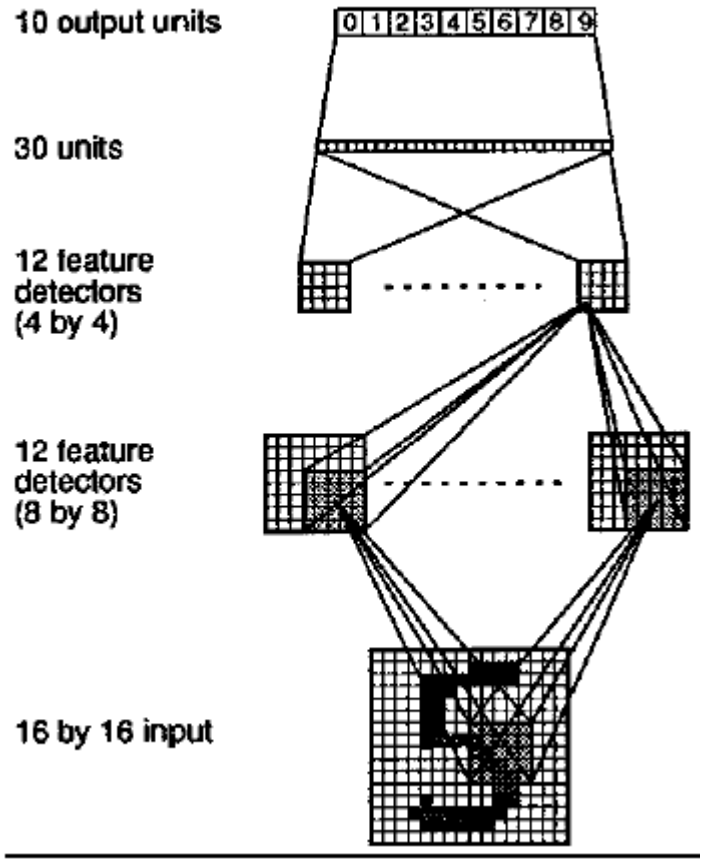


Normalize Digits First

Feature Detectors



Network Structure



Atypical Data Recognized

Further Details and Results

- ~10,000 digits from the U.S. mail were used to train and test system
- ZIP codes on envelopes were initially located and segmented by a separate system (difficult task in itself)
- *weight sharing* used to constrain number of free parameters
- 1,256 units + 30,060 links + 1,000 biases, but only 9760 free parameters
- used an accelerated version of backprop (pseudo-Newton rule)
- trained on 7,300 digits, tested on 2,000
- error rate of ~1% on training set, ~5% on test set
- if marginal cases were rejected (two or more outputs approximately the same), then error reduced to ~1% with 12% rejected
- used "optimal brain damage" technique to prune unnecessary weights
- after removing weights and retraining, only ~1/4 as many free parameters as before, but better performance: 99% accuracy with 9% rejection rate
- achieved state-of-the-art in digit recognition
- much problem-specific knowledge was put into the network architecture
- preprocessing of input data was crucial to success

ALVINN

(Autonomous Land Vehicle In a Neural Network)

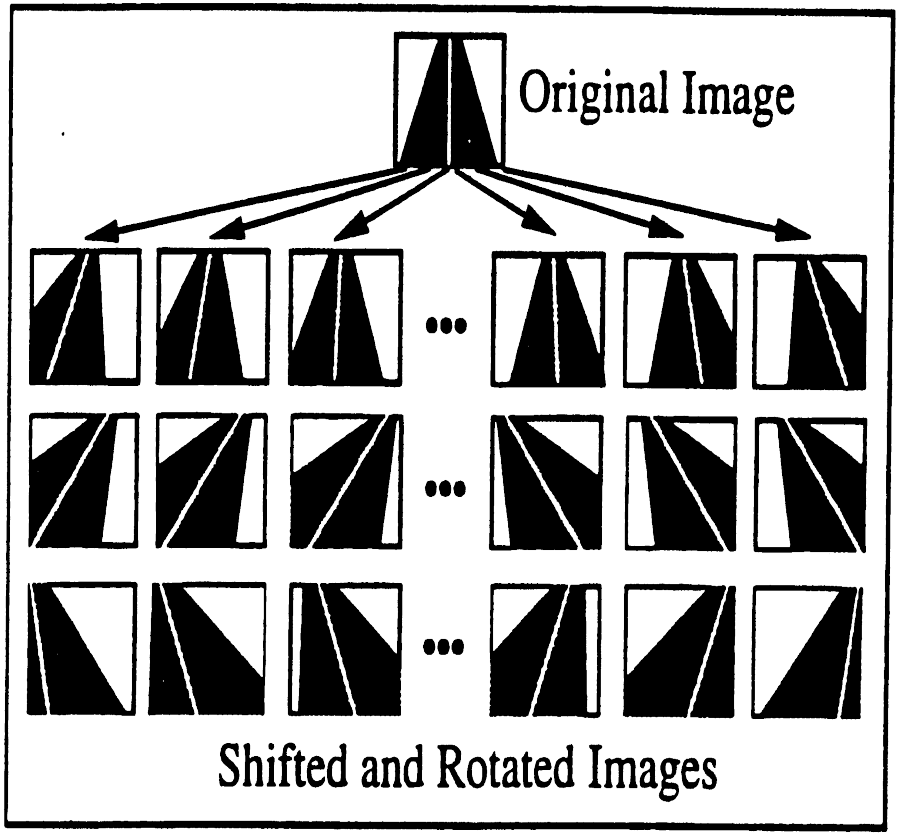
(Pomerleau, 1996)

- Network-controlled steering of a car on a winding road
- network inputs: 30 x 32 pixel image from a video camera, 8 x 32 gray scale image from
- a range finder
- 29 hidden units
- 45 output units arranged in a line corresponding to steering angle
- achieved speeds of up to 70 mph for 90 minutes on highways outside of Pittsburgh

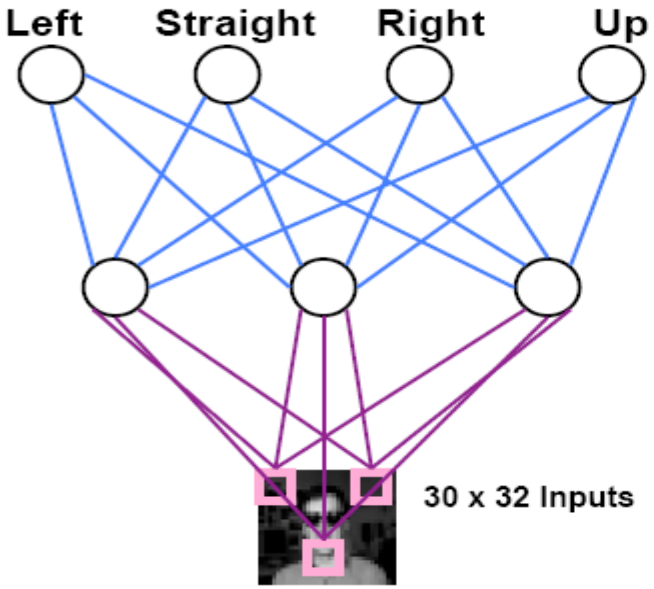
ALVINN – Enhancing Training

Training set collected by having a human drive the vehicle: the human is too good!

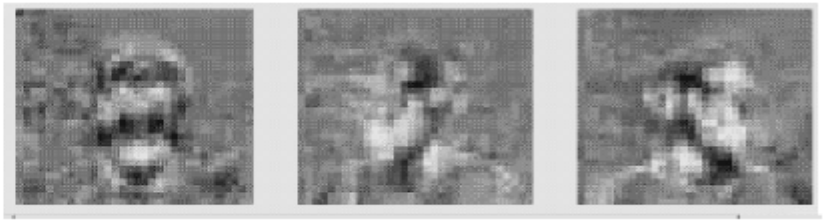
Solution: Rotating each image to create additional views

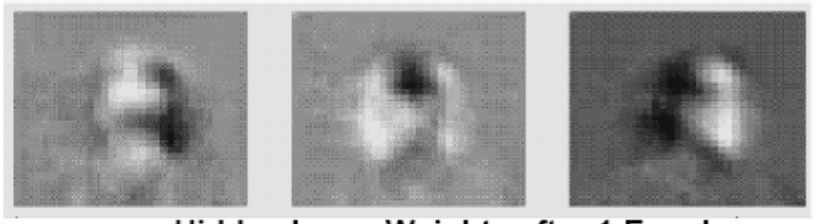


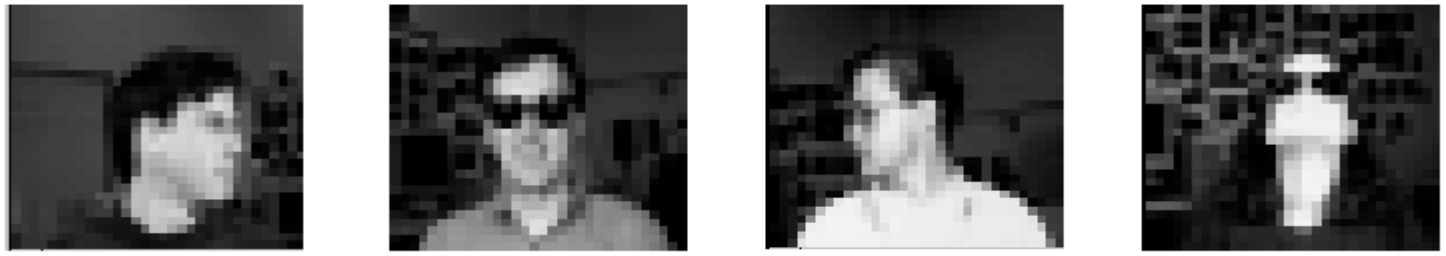
Face Recognition (Mitchell, 1997)




Output Layer Weights (including $w_0 = \theta$) after 1 Epoch


Hidden Layer Weights after 25 Epochs


Hidden Layer Weights after 1 Epoch



- 90% Accurate Learning Head Pose, recognizing 1-of-20 Faces (more info at http://www.cs.cmu.edu/~tom/faces.html)

Some additional examples

application	#weights	#samples	error	ref.
text -> speech	25000	5000	0.20	Sejnowski
sonar target rec	1105	192	0.15	Gorman
car control	>36000	1200	car drives on winding road	Pomerleau
back-gammon	>11000	3000	computer champion	Tesauro
sex rec from faces	>36000	90	0.09	Golomb
char rec	9900	5000	0.055	Sato
remote sensing	1800	50	0.05-0.10	Kamata
signature verif.	480	280	0.05	Sabourin

T.J. Sejnowski and C.R. Rosenberg, *NETtalk: a parallel network that learns to read aloud*, The John Hopkins University Electrical Eng. and Comp. Science, 1986.

P. Gorman and T.J. Sejnowski, *Learned Classification of Sonar Targets Using Massively Parallel Network*, IEEE Transactions on ASSP, vol. 36, no. 7, July 1988.

D. Pomerleau, ALVINN: *Ann Autonomous Land Vehicle in a Neural Network*, in: David S. Touretzky, Advances in Neural Information Processing Systems I, 1989

G. Tesauro, *Neurogammon wins computer olympiad*, Neural Computation, vol. 1, pp 312-323, 1990

B.A. Golomb, D.T. Lawrence, T.J. Sejnowski, Sexnet: *A neural network identifies sex from human faces*, Adv. in Neural Inf. Proc. Sys. I, 1989

A. Sato, K. Yamada, J. Tsukumo, and T. Temma, *Neural network models for incremental learning*, ICNN, Helsinki, 1991.

S.-I. Kamata, R.O. Eason, A. Perez, and E. Kawaguchi, *A Neural Network Classifier for LANDSAT Image Data*, Proc. 11th ICPR, The Hague, Vol 2, 573-576, 1992

R. Sabourin and J-P. Drouhard, *Off-Line Signature Verification Using Directional PDF and Neural Networks*, Proc. 11th ICPR, The Hague, Vol 2, 321-325, 1992