Unit – IV

Application of Neural Network

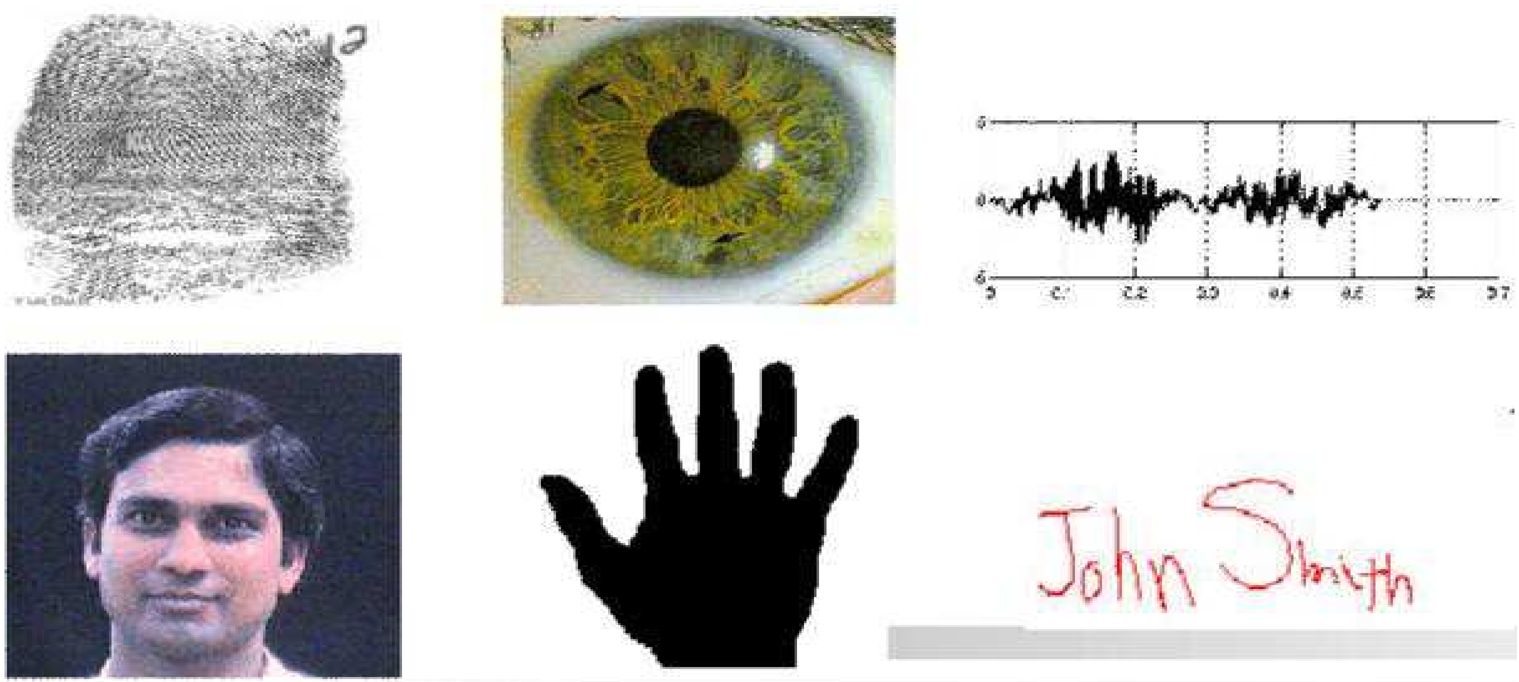
What is Pattern Recognition (PR)?

- It is the study of how machines can
  - observe the environment
  - learn to distinguish patterns of interest from their background
  - make sound and reasonable decisions about the categories of the patterns.

What is a pattern?

- Watanable [163] defines a pattern as

“ *the opposite of a chaos; it is an entity, vaguely defined, that could be given a name*”



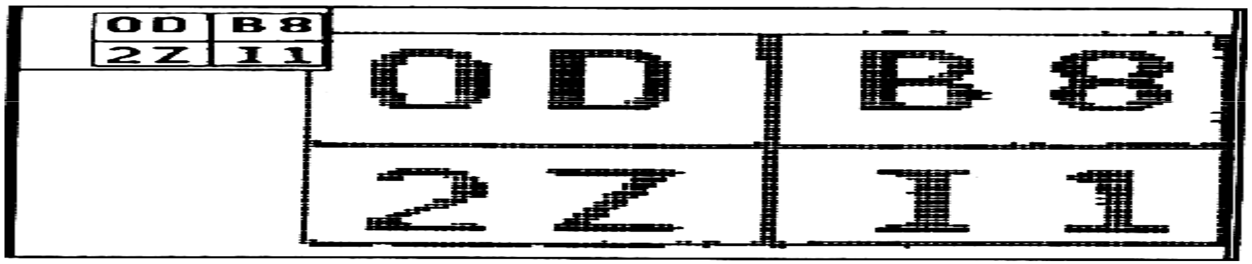
Other Patterns

- Insurance, credit card applications - applicants are characterized by
  - # of accidents, make of car, year of model
  - Income, # of dependents, credit worthiness, mortgage amount
- Dating services
  - Age, hobbies, income, etc. establish your “desirability”
- Web documents
  - Key words based descriptions (e.g., documents containing “terrorism”, “Osama” are different from those containing “football”, “NFL”).
- Housing market
  - Location, size, year, school district, etc.

Pattern Class

- A collection of "similar" (not necessarily identical) objects
  - Inter-class variability

  - Intra-class variability

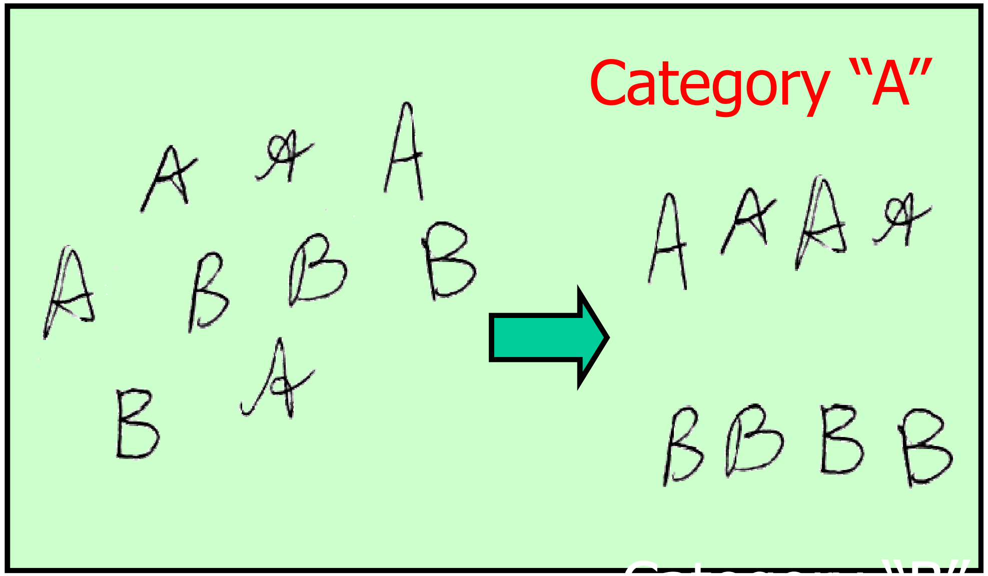


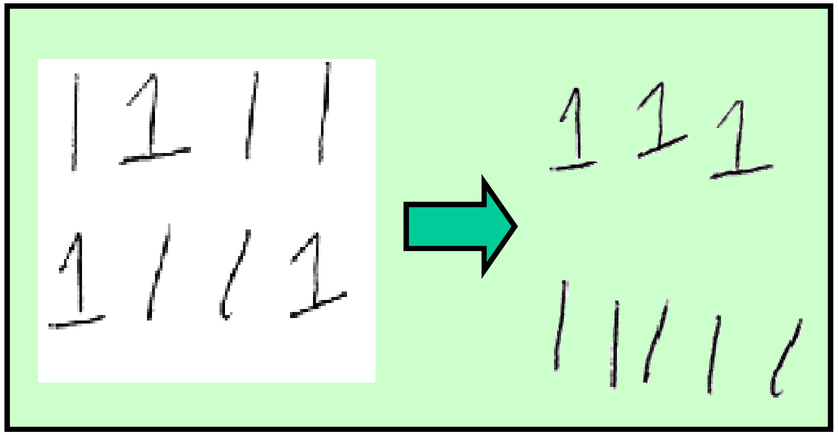
Pattern Class Model

- Different descriptions, which are typically mathematical in form for each class/population (e.g., a probability density like Gaussian)

Classification vs Clustering

- Classification (known categories)
- Clustering (creation of new categories)





(Unsupervised Classification)

(Supervised Classification)

Pattern Recognition

- Key Objectives:
  - Process the sensed data to eliminate noise
  - Hypothesize the models that describe each class population (e.g., recover the process that generated the patterns).
  - Given a sensed pattern, choose the best-fitting model for it and then assign it to class associated with the model.

Emerging PR Applications

Problem	Input	Output
Speech recognition	Speech waveforms	Spoken words, speaker identity
Non-destructive testing	Ultrasound, eddy current, acoustic emission waveforms	Presence/absence of flaw, type of flaw
Detection and diagnosis of disease	EKG, EEG waveforms	Types of cardiac conditions, classes of brain conditions
Natural resource identification	Multispectral images	Terrain forms, vegetation cover
Aerial reconnaissance	Visual, infrared, radar images	Tanks, airfields
Character recognition (page readers, zip code, license plate)	Optical scanned image	Alphanumeric characters

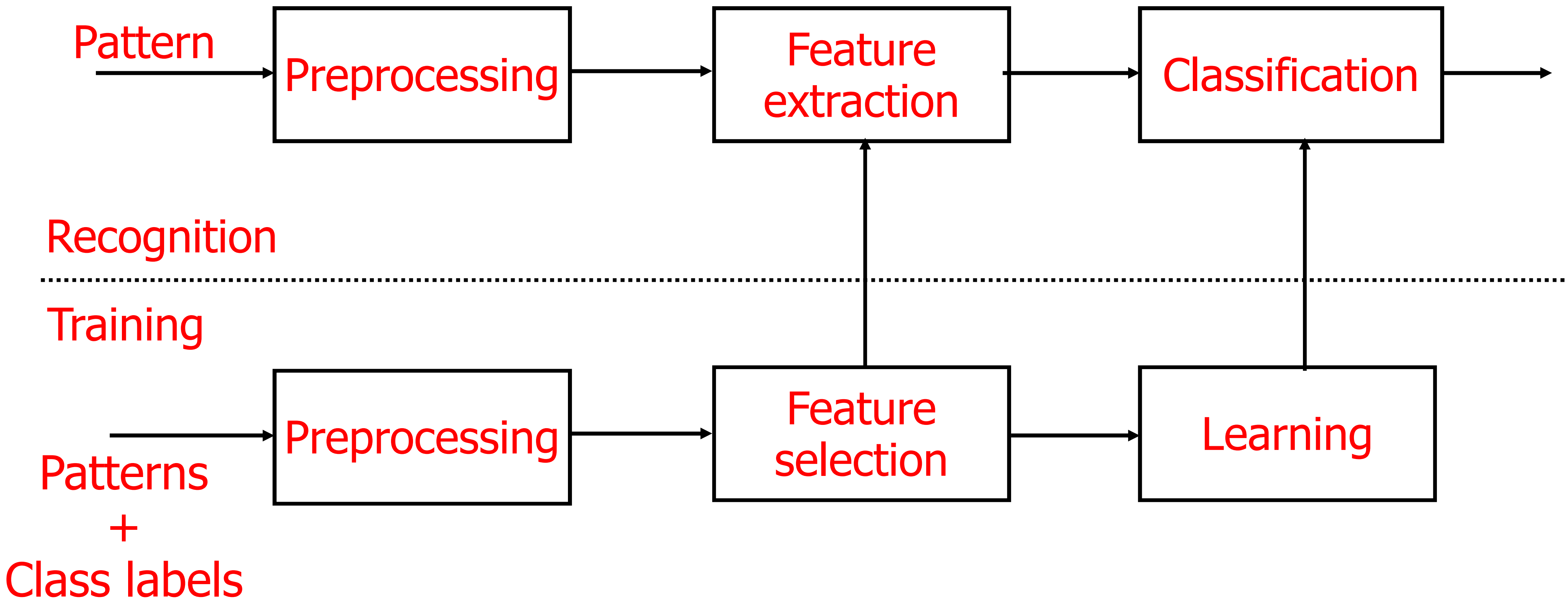
Emerging PR Applications (cont'd)

Problem	Input	Output
Identification and counting of cells	Slides of blood samples, micro-sections of tissues	Type of cells
Inspection (PC boards, IC masks, textiles)	Scanned image (visible, infrared)	Acceptable/unacceptable
Manufacturing	3-D images (structured light, laser, stereo)	Identify objects, pose, assembly
Web search	Key words specified by a user	Text relevant to the user
Fingerprint identification	Input image from fingerprint sensors	Owner of the fingerprint, fingerprint classes
Online handwriting retrieval	Query word written by a user	Occurrence of the word in the database

Main PR Areas

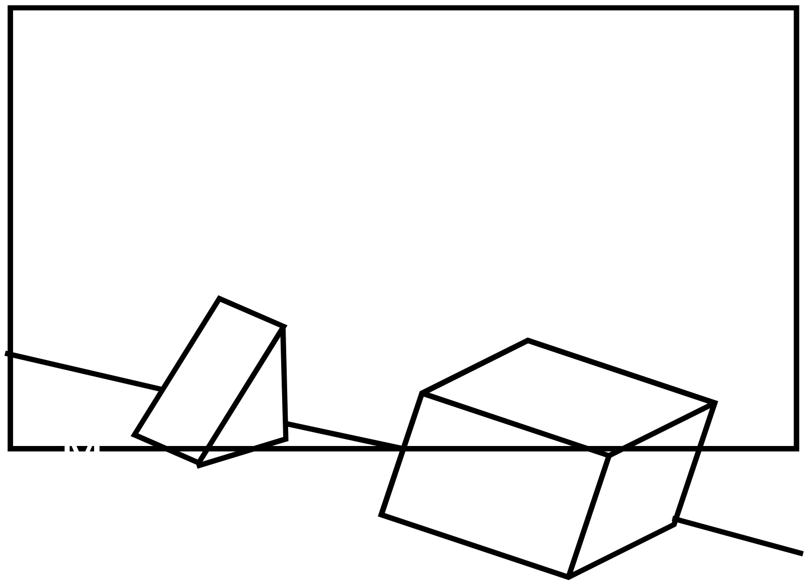
- *Template matching*
  - The pattern to be recognized is matched against a stored template while taking into account all allowable pose (translation and rotation) and scale changes.
- *Statistical pattern recognition*
  - Focuses on the statistical properties of the patterns (i.e., probability densities).
- *Structural Pattern Recognition*
  - Describe complicated objects in terms of simple primitives and structural relationships.
- *Syntactic pattern recognition*
  - Decisions consist of logical rules or grammars.
- *Artificial Neural Networks*
  - Inspired by biological neural network models.

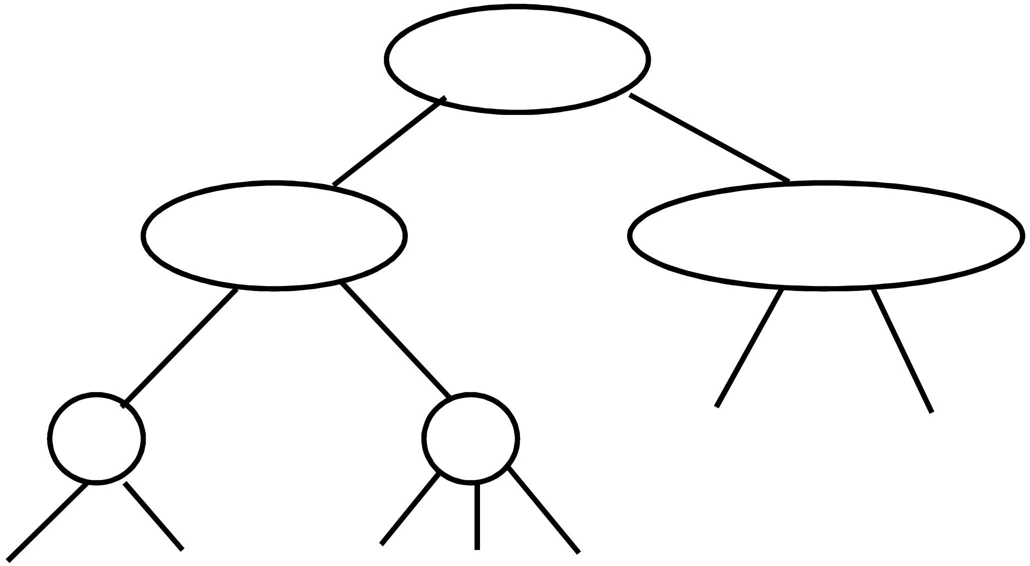
Statistical Pattern Recognition



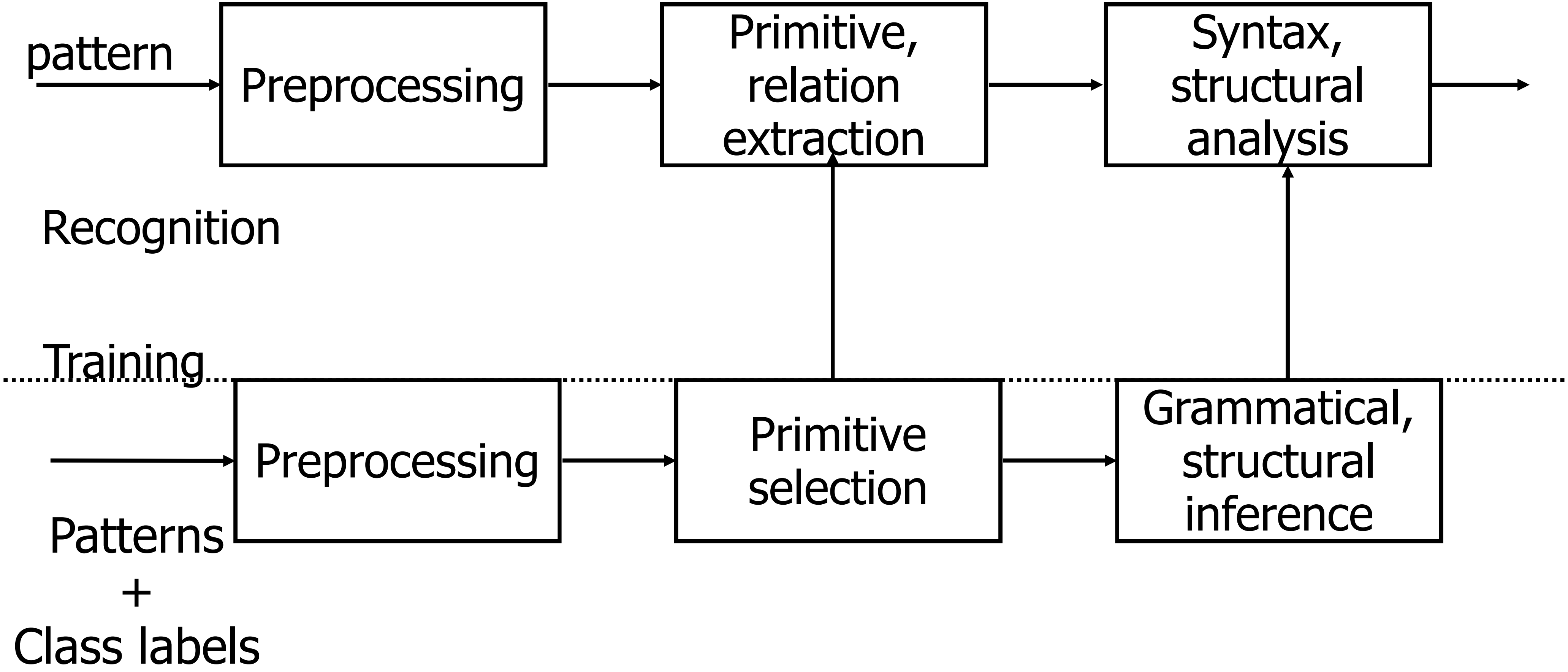
Structural Pattern Recognition

- Describe complicated objects in terms of simple primitives and structural relationships.
- Decision-making when features are non-numeric or structural



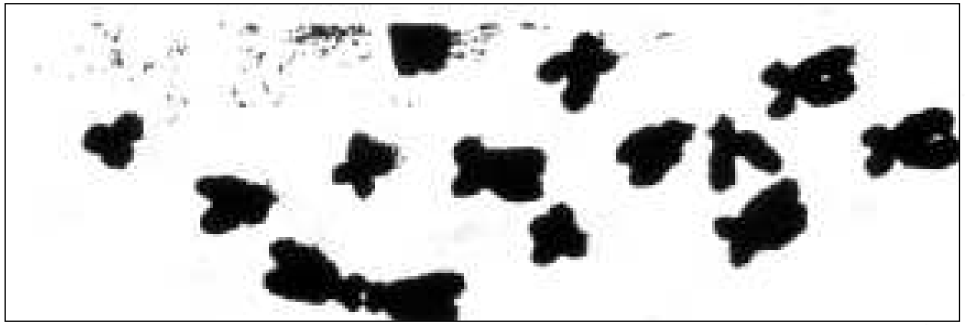


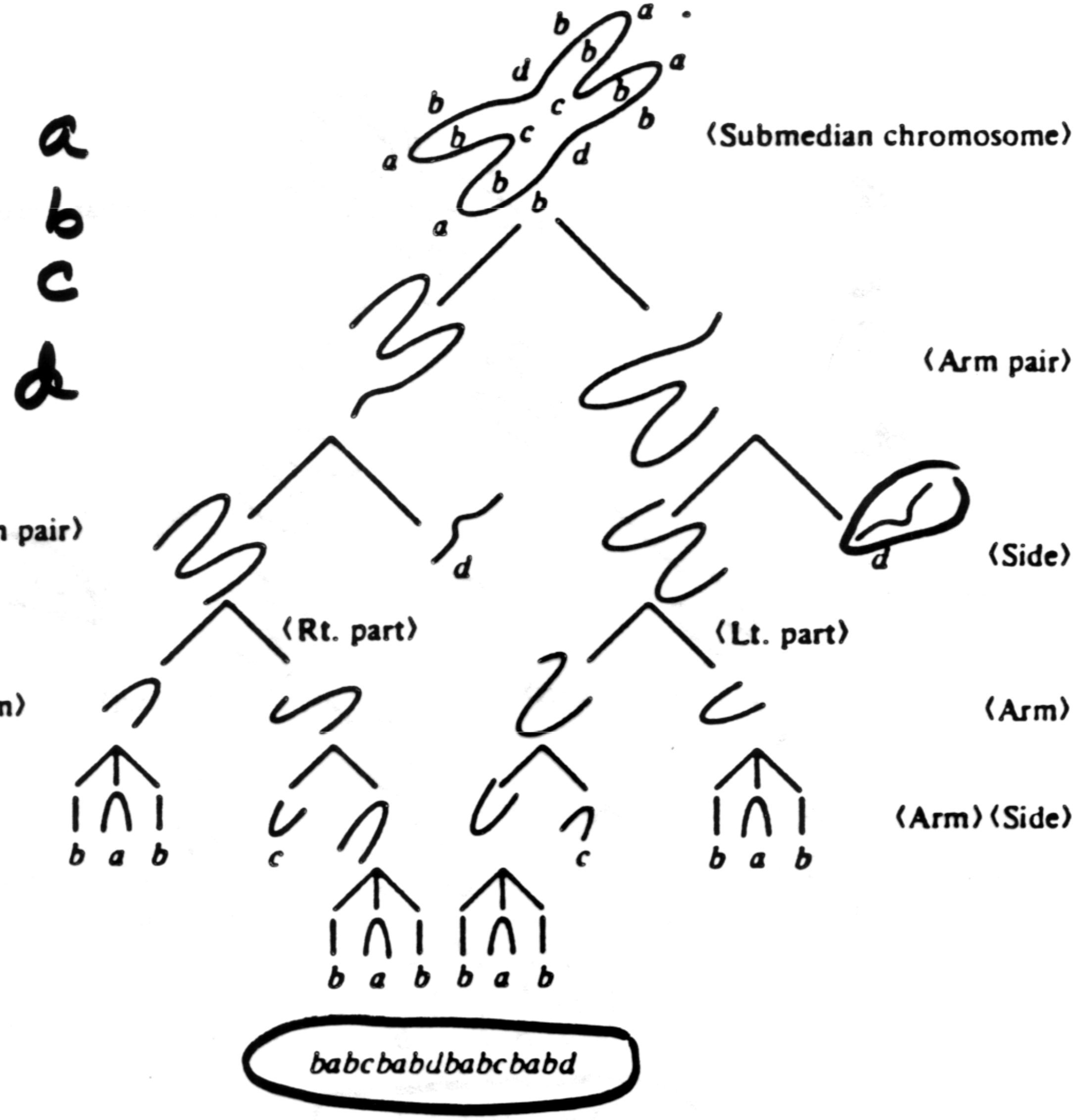
Syntactic Pattern Recognition



Describe patterns using deterministic grammars or formal languages

Chromosome Grammars



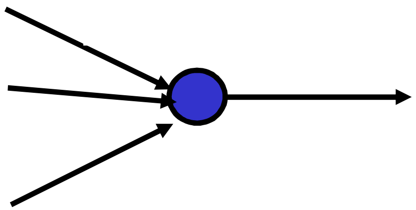


Artificial Neural Networks

- Massive parallelism is essential for complex pattern recognition tasks (e.g., speech and image recognition)
  - Human take only a few hundred ms for most cognitive tasks; suggests parallel computation
- Biological networks attempt to achieve good performance via dense interconnection of simple computational elements (neurons)
  - Number of neurons $\approx 10^{10} – 10^{12}$
  - Number of interconnections/neuron $\approx 10^{3} – 10^{4}$
  - Total number of interconnections $\approx 10^{14}$

Artificial Neural Nodes

- Nodes in neural networks are nonlinear, typically analog

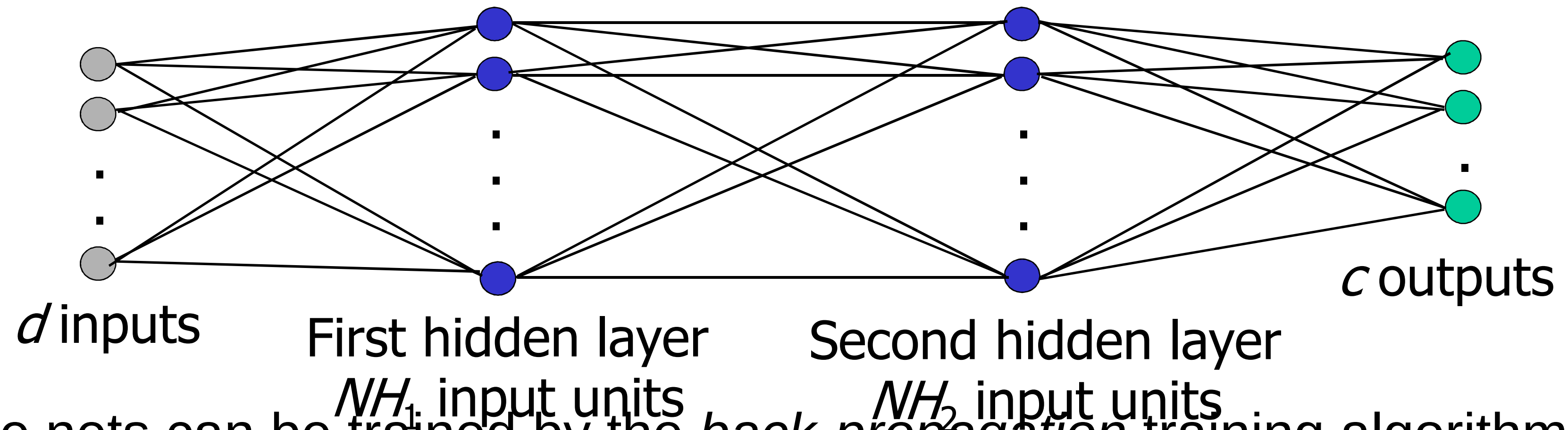


$$Y = f\left(\sum_{i=1}^{d} w_i x_i - \theta\right)$$

where θ is an internal threshold

Multilayer Perceptron

- Feed-forward nets with one or more layers (hidden) between the input and output nodes
- A three-layer net can generate arbitrary complex decision regions



- These nets can be trained by the *back-propagation* training algorithm

Comparing Pattern Recognition Models

- Template Matching
  - Assumes very small intra-class variability
  - Learning is difficult for deformable templates
- Structural / Syntactic
  - Primitive extraction is sensitive to noise
  - Describing a pattern in terms of primitives is difficult
- Statistical
  - Assumption of density model for each class
- Artificial Neural Network
  - Parameter tuning and local minima in learning

Speech Recognition

Definition

- Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words.
- The recognised words can be an end in themselves, as for applications such as commands & control, data entry, and document preparation.
- They can also serve as the input to further linguistic processing in order to achieve speech understanding

Speech Processing

- Signal processing:
  - Convert the audio wave into a sequence of feature vectors
- Speech recognition:
  - Decode the sequence of feature vectors into a sequence of words
- Semantic interpretation:
  - Determine the meaning of the recognized words
- Dialog Management:
  - Correct errors and help get the task done
- Response Generation
  - What words to use to maximize user understanding
- Speech synthesis (Text to Speech):
  - Generate synthetic speech from a ‘marked-up’ word string

What you can do with Speech Recognition

- Transcription
  - dictation, information retrieval
- Command and control
  - data entry, device control, navigation, call routing
- Information access
  - airline schedules, stock quotes, directory assistance
- Problem solving
  - travel planning, logistics

Transcription and Dictation

- Transcription is transforming a stream of human speech into computer-readable form
  - Medical reports, court proceedings, notes
  - Indexing (e.g., broadcasts)
- Dictation is the interactive composition of text
  - Report, correspondence, etc.

Speech recognition and understanding

- Sphinx system
  - speaker-independent
  - continuous speech
  - large vocabulary
- ATIS system
  - air travel information retrieval
  - context management

Speech Recognition and Call Centres

- Automate services, lower payroll
- Shorten time on hold
- Shorten agent and client call time
- Reduce fraud
- Improve customer service

Applications related to Speech Recognition

- Speech Recognition
- Figure out what a person is saying.
- Speaker Verification
- Authenticate that a person is who she/he claims to be.
- Limited speech patterns
- Speaker Identification
- Assigns an identity to the voice of an unknown person.
- Arbitrary speech patterns

kinds of Speech Recognition Systems

- Speech recognition systems can be characterised by many parameters.
- An isolated-word (**Discrete**) speech recognition system requires that the speaker pauses briefly between words, whereas a **continuous** speech recognition system does not.

A TIMELINE OF SPEECH RECOGNITION

- **1890s Alexander Graham Bell discovers Phone while trying to develop speech recognition system for deaf people.**
- **1936**AT&T's Bell Labs produced the first electronic speech synthesizer called the Voder (Dudley, Riesz and Watkins).
- This machine was demonstrated in the 1939 World Fairs by experts that used a keyboard and foot pedals to play the machine and emit speech.
- **1969**John Pierce of Bell Labs said automatic speech recognition will not be a reality for several decades because it requires artificial intelligence.

Early 70s

- **Early 1970's**The Hidden Markov Modeling (HMM) approach to speech recognition was invented by Lenny Baum of Princeton University and shared with several ARPA (Advanced Research Projects Agency) contractors including IBM.
- HMM is a complex mathematical pattern-matching strategy that eventually was adopted by all the leading speech recognition companies including Dragon Systems, IBM, Philips, AT&T and others.

70+

- **1971**DARPA (Defense Advanced Research Projects Agency) established the Speech Understanding Research (SUR) program to develop a computer system that could understand continuous speech.
- Lawrence Roberts, who initiated the program, spent $3 million per year of government funds for 5 years. Major SUR project groups were established at CMU, SRI, MIT's Lincoln Laboratory, Systems Development Corporation (SDC), and Bolt, Beranek, and Newman (BBN). It was the largest speech recognition project ever.
- **1978**The popular toy "Speak and Spell" by Texas Instruments was introduced. Speak and Spell used a speech chip which led to huge strides in development of more human-like digital synthesis sound.

80+

- **1982**Covox founded. Company brought digital sound (via The Voice Master, Sound Master and The Speech Thing) to the Commodore 64, Atari 400/800, and finally to the IBM PC in the mid ‘80s.
- **1982**Dragon Systems was founded in 1982 by speech industry pioneers Drs. Jim and Janet Baker. Dragon Systems is well known for its long history of speech and language technology innovations and its large patent portfolio.
- **1984**SpeechWorks, the leading provider of over-the-telephone automated speech recognition (ASR) solutions, was founded.

90s

- **1993** Covox sells its products out to Creative Labs, Inc.
- **1995** Dragon released discrete word dictation-level speech recognition software. It was the first time dictation speech recognition technology was available to consumers. IBM and Kurzweil followed a few months later.
- **1996** Charles Schwab is the first company to devote resources towards developing up a speech recognition IVR system with Nuance. The program, Voice Broker, allows for up to 360 simultaneous customers to call in and get quotes on stock and options... it handles up to 50,000 requests each day. The system was found to be 95% accurate and set the stage for other companies such as Sears, Roebuck and Co., and United Parcel Service of America Inc., and E*Trade Securities to follow in their footsteps.
- **1996** BellSouth launches the world's first voice portal, called Val and later Info By Voice.

95+

- **1997** Dragon introduced "Naturally Speaking", the first "continuous speech" dictation software available (meaning you no longer need to pause between words for the computer to understand what you're saying).
- **1998** Lernout & Hauspie bought Kurzweil. Microsoft invested $45 million in Lernout & Hauspie to form a partnership that will eventually allow Microsoft to use their speech recognition technology in their systems.
- **1999** Microsoft acquired Entropic, giving Microsoft access to what was known as the "most accurate speech recognition system" in the Old VCR!

2000

2000 Lernout & Hauspie acquired Dragon Systems for approximately $460 million.

2000 TellMe introduces first world-wide voice portal.

2000 NetBytel launched the world's first voice enabler, which includes an on-line ordering application with real-time Internet integration for Office Depot.

2000s

2001 ScanSoft Closes Acquisition of Lernout & Hauspie Speech and Language Assets.

2003 ScanSoft Ships Dragon NaturallySpeaking 7 Medical, Lowers Healthcare Costs through Highly Accurate Speech Recognition.

2003 ScanSoft closes deal to distribute and support IBM ViaVoice Desktop Products.

Signal Variability

- Speech recognition is a difficult problem, largely because of the many sources of variability associated with the signal.
- The acoustic realisations of phonemes, the recognition systems smallest sound units of which words are composed, are highly dependent on the context in which they appear.
- These phonetic variables are exemplified by the acoustic differences of the phoneme 't/'in two, true, and butter in English.
- At word boundaries, contextual variations can be quite dramatic, and devo andare sound like devandare in Italian.

What is a speech recognition system?

- Speech recognition is generally used as a human computer interface for other software. When it functions in this role, three primary tasks need be performed.
- Pre-processing, the conversion of spoken input into a form the recogniser can process.
- Recognition, the identification of what has been said.
- Communication, to send the recognised input to the application that requested it.

How is pre-processing performed

- To understand how the first of these functions is performed, we must examine,
- Articulation, the production of the sound.
- Acoustics, the stream of the speech itself.
- What characterises the ability to understand spoke input, Auditory perception.

Neural networks

- "if speech recognition systems could learn speech knowledge automatically and represent this knowledge in a parallel distributed fashion for rapid evaluation … such a system would mimic the function of the human brain, which consists of several billion simple, inaccurate and slow processors that perform reliable speech processing", (Waibel and Hampshire, 1989).
- An artificial neural network is a computer program, which attempt to emulate the biological functions of the Human brain. They are an excellent classification systems, and have been effective with noisy, patterned, variable data streams containing multiple, overlapping, interacting and incomplete cues, (Markowitz, 1995).

Neural networks

- Neural networks do not require the complete specification of a problem, learning instead through exposure to large amount of example data. Neural networks comprise of an input layer, one or more hidden layers, and one output layer. The way in which the nodes and layers of a network are organised is called the networks architecture.
- The allure of neural networks for speech recognition lies in their superior classification abilities.
- Considerable effort has been directed towards development of networks to do word, syllable and phoneme classification.

The Phonetic Typewriter

Developed for Finnish (a phonetic language, written as it is said).
Trained on one speaker, will generalise to others.

A neural network is trained to cluster together similar sounds, which are then labelled with the corresponding character.

When recognising speech, the sounds uttered are allocated to the closest corresponding output, and the character for that output is printed.

- requires large dictionary of minor variations to correct general mechanism

- noticeably poorer performance on speakers it has not been trained on

The Phonetic Typewriter (cont'd)

a	a	a	ah	h	æ	æ	ø	ø	e	e	e
o	a	a	h	r	æ	l	ø	y	y	j	i
o	o	a	h	r	r	r	g	g	y	j	i
o	o	m	a	r	m	n	m	n	j	i	i
l	o	u	h	v	vm	n	n	h	hj	j	j
l	u	v	v	p	d	d	t	r	h	hi	j
.	.	u	v	tk	k	p	p	p	r	k	s
.	.	v	k	pt	t	p	t	p	h	s	s

Character recognition Neural networks

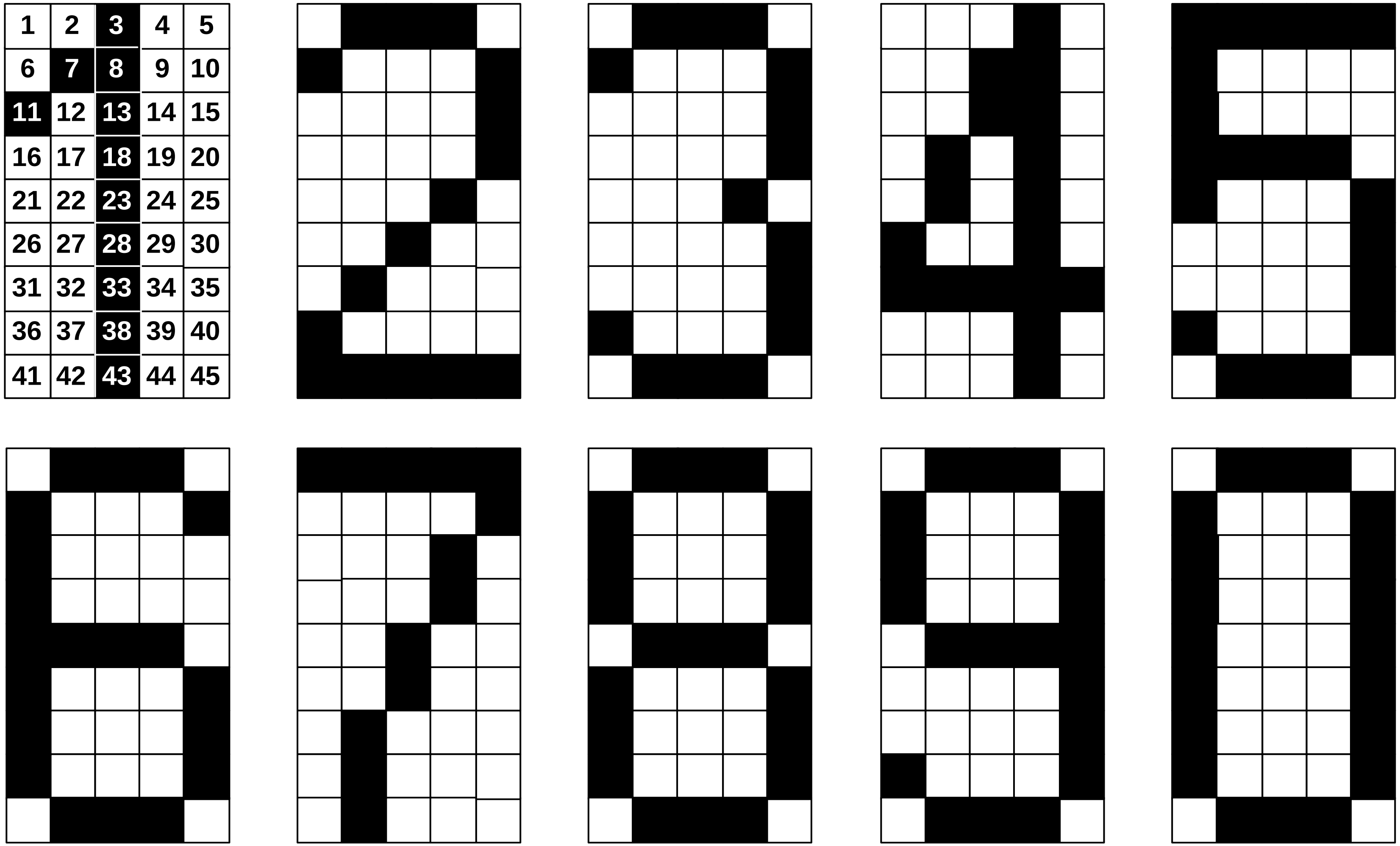
Recognition of both printed and handwritten characters is a typical domain where neural networks have been successfully applied.

A multilayer feedforward network for printed character recognition can be used.

For simplicity, you can limit your task to the recognition of digits from 0 to 9. Each digit is represented by a 5 ´ 9 bit map

In commercial applications, where a better resolution is required, at least 16 ´ 16 bit maps are used.

Bit maps for digit recognition



How do choose the architecture of a neural network?

- The number of neurons in the input layer is decided by the number of pixels in the bit map. The bit map in our example consists of 45 pixels, and thus we need 45 input neurons.

- The output layer has 10 neurons – one neuron for each digit to be recognized.

How do we determine an optimal number of hidden neurons?

- Complex patterns cannot be detected by a small number of hidden neurons; however too many of them can dramatically increase the computational burden.
- Another problem is **overfitting** The greater the number of hidden neurons, the greater the ability of the network to recognise existing patterns. However, if the number of hidden neurons is too big, the network might simply memorise all training examples.

Neural network for printed digit recognition