MACHINE LEARNING

Science is like sex: sometimes something useful
come out, but that is not the reason we are doing it

– Richard Feynman

Marco Gori
Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche

Università di Siena

Machine Learning, University of Siena
2018-2019

Machine Learning

A CONSTRAINT-BASED APPROACH

Marco Gori

Contents

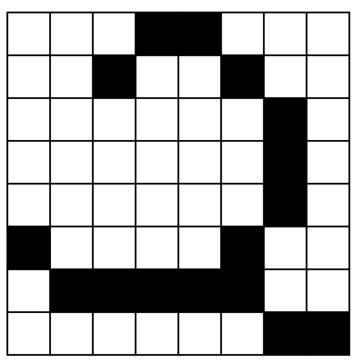
CHAPTER

The Big Picture

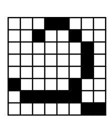
1

Let's start!

Why do machines need to learn?



Handwritten characters: The 2^d warning!

 $\sim$ 0001100000100100000000100000001000000010100001000111110000000011.

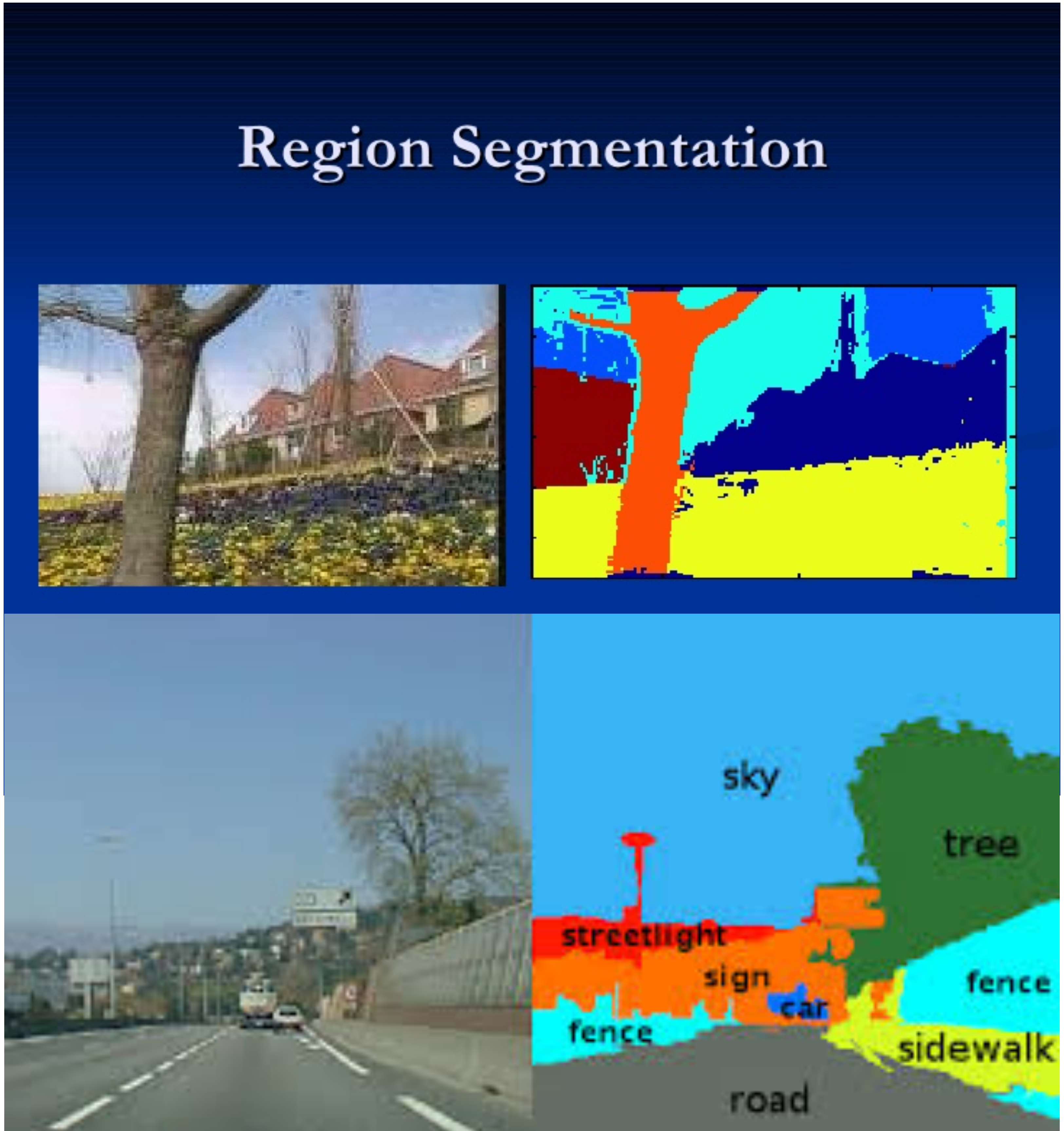
Patterns and segmentation

Segmentation might be as difficult as recognition!

signal. Unfortunately, those analyses are doomed to fail. The sentence "`computers are attacking the secret of intelligence`", quickly pronounced, would likely

`com / pu / tersarea / tta / ckingthesecre / tofin / telligence`.

In vision

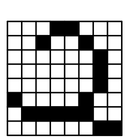

Region Segmentation
sky
tree
streetlight
sign
car
fence
fence
sidewalk
road

Learning tasks

Agent: $\chi : \mathscr{E} \to \mathscr{D}$.

$$\pi : \mathscr{E} \to \mathscr{X}$$
$$f : \mathscr{X} \to \mathscr{Y}$$
$$h : \mathscr{Y} \to \mathscr{O}$$

$$\chi = h \circ f \circ \pi,$$
where π *is the input encoding,* f *is the learning function, and* h *is the output encoding.*

 $\xrightarrow{\pi} (0, 0, 0, 1, 1, 0, 0, 0, \ldots, 0, 0, 0, 0, 0, 0, 1, 1)'$

$\xrightarrow{f} (0, 0, 1, 0, 0, 0, 0, 0, 0, 0)' \xrightarrow{h} 2.$

Regression and classification

Agent: $\chi : \mathscr{E} \to \mathscr{D}$.

$$\mathscr{O} \subset \mathbb{N}$$

$$|\mathscr{O}| = 10.$$

$$\mathscr{O} = \mathbb{R}$$

Structured representations

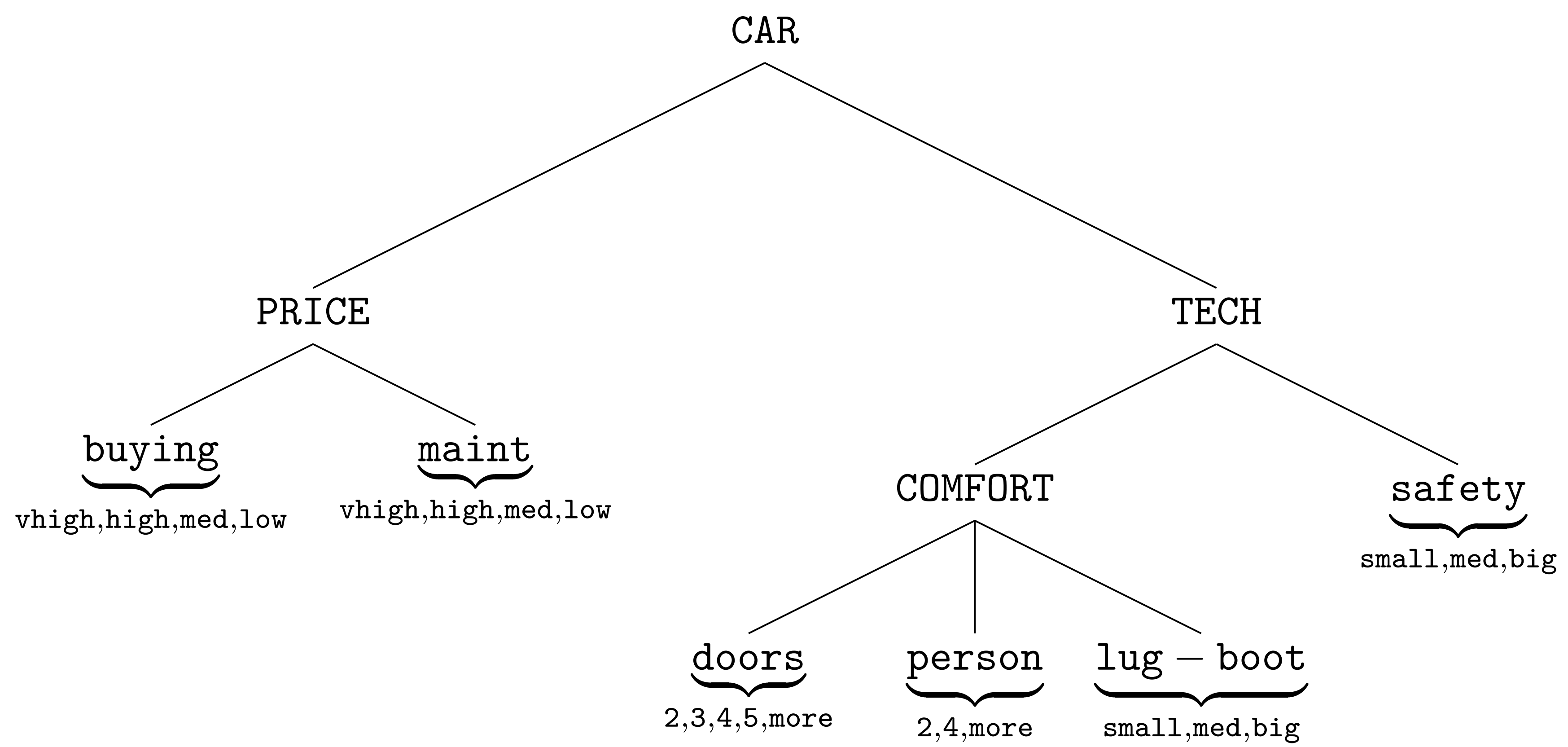


FIGURE 1.1

This learning task is presented in the UCI Machine Learning repository https://archive.ics.uci.edu/ml/datasets/Car+Evaluation.

Structured representations (con't)

(A) (B)

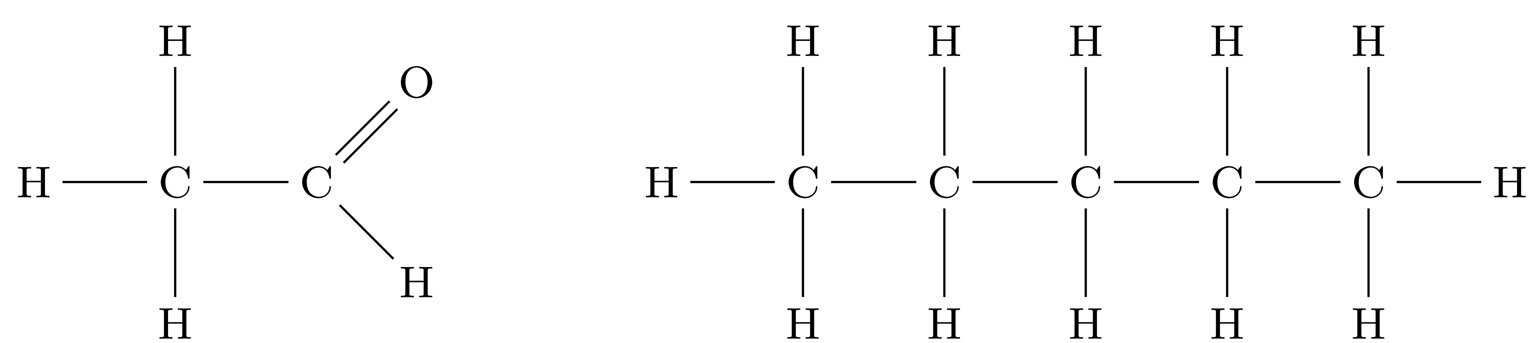


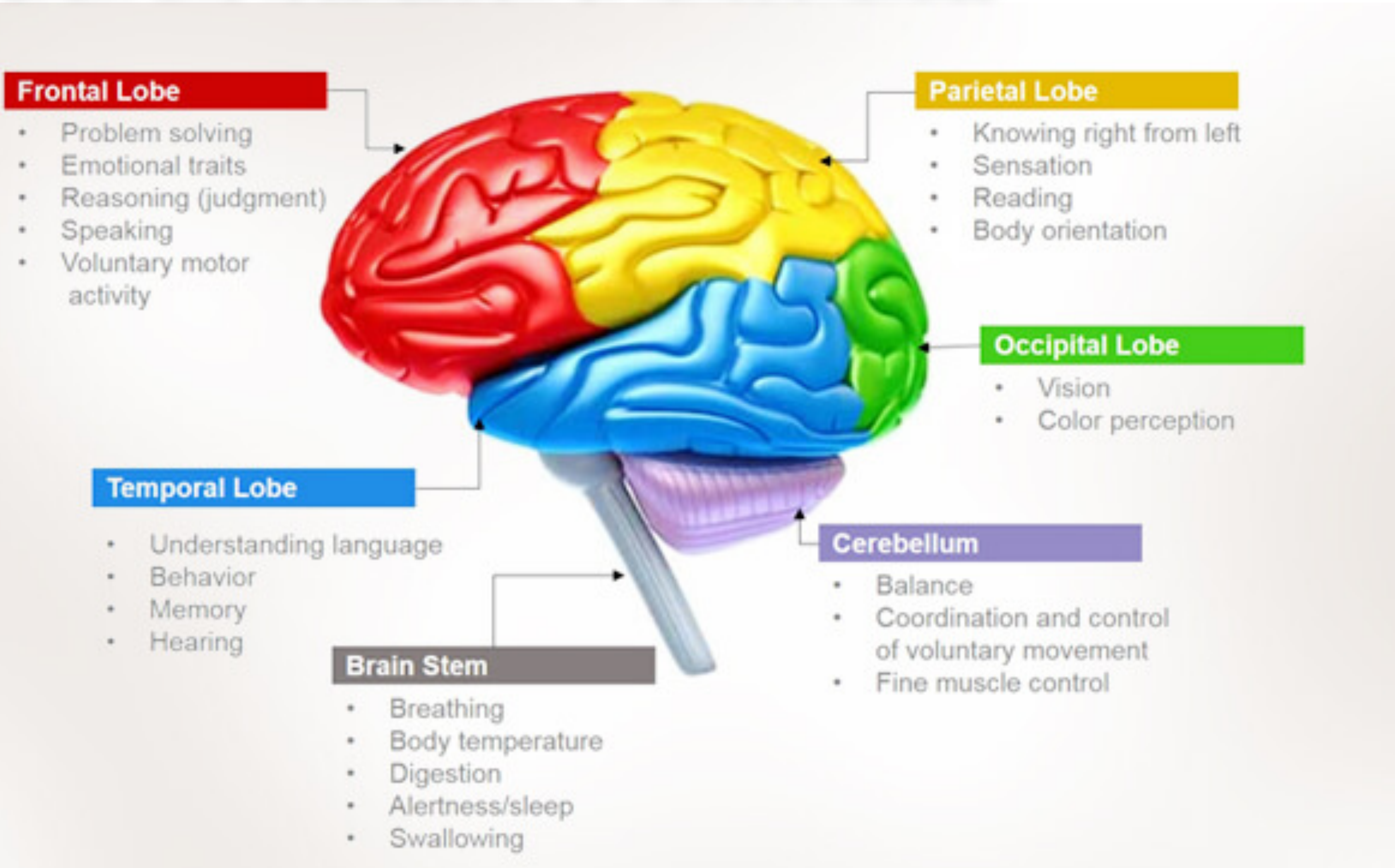
FIGURE 1.2

Two chemical formulas: (A) acetaldehyde with formula CH_3CHO, (B) N-heptane with the chemical formula $H_3C(CH_2)_5CH_3$.

Biological neurons

about 100 billion, 7,000 synaptic connections each

Brain and localization of functionalities



Artificial neurons

$$a_i = b_i + \sum_{j=1}^{d} w_{i,j} x_j,$$

$$y_i = \sigma(a_i) = 1/(1 + e^{-a_i}).$$

Feedforward neural network

$$\mathrm{f}_j(w, x_\kappa)$$

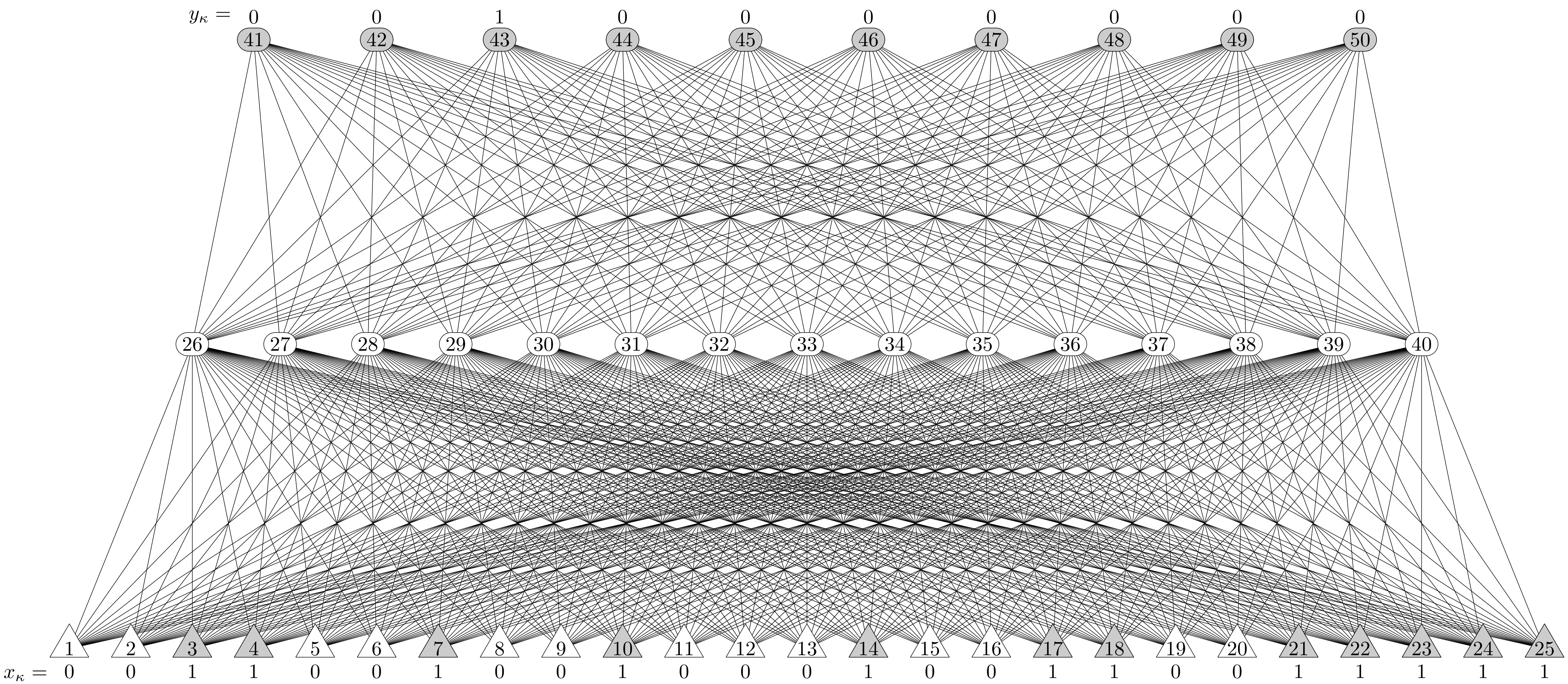


FIGURE 1.3

Recognition of handwritten chars. The incoming pattern $x = \pi(\cdot)$ is processed by the feedforward neural network, whose target consists of firing only neuron 43. This corresponds with the *one-hot* encoding of class "2".

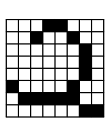
LEARNING PROTOCOLS

Supervised learning

$$\mathscr{L} = \{(e_1, o_1), \ldots, (e_\ell, o_\ell)\}$$

2

$$\{(x_\kappa, y_\kappa),\ \kappa = 1, \ldots, \ell\}$$

 $\sim$ 000110000010010000000001000000010000000101000010001111100000000011.

Error function

$$(\mathscr{N}, \mathscr{L}) \rightsquigarrow E(\cdot)$$

$$E(w) = \sum_{\kappa=1}^{\ell} \sum_{j=1}^{n} (1 - y_{\kappa j} \mathsf{f}_j(w, x_\kappa))_+$$

Unsupervised learning

$$x \in \mathscr{X} \subset \mathbb{R}^d$$

$$\overline{x} \in \mathscr{X}$$

$$\|x - \overline{x}\| < \rho$$

$$\mathscr{N}_\rho = \{ x \in \mathscr{X} \mid \|x - \overline{x}\| < \rho \}$$

$$\mathrm{vol}\,(\mathscr{N}_\rho) = \frac{(\sqrt{\pi})^d}{\Gamma(1 + \frac{d}{2})} \rho^d$$

Everything is in the peel!

Space oddities at high dimensions.

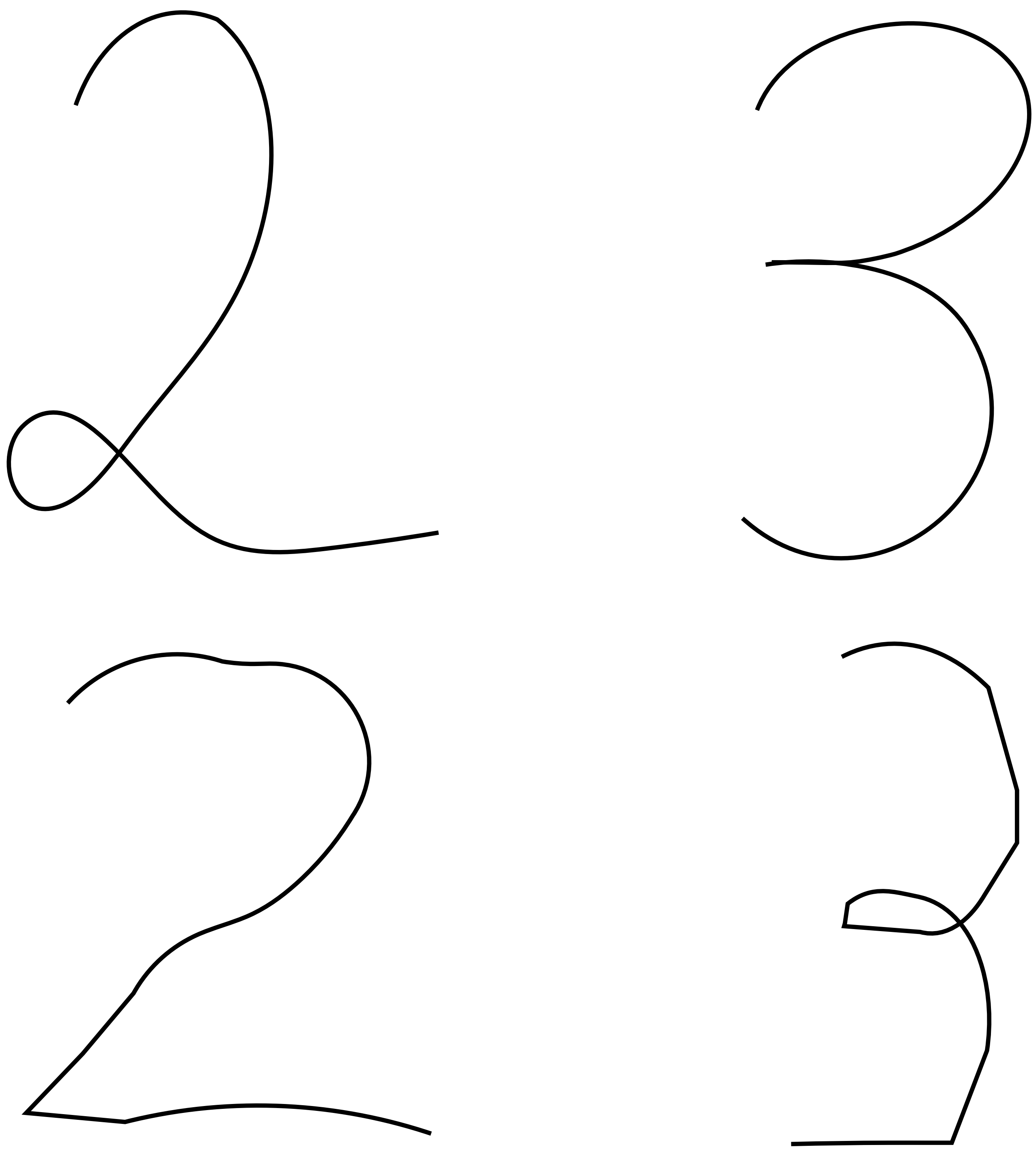
$$\mathscr{P}_\epsilon = \{x \in \mathscr{X} \mid \|x - \overline{x}\| < \rho \quad \text{and} \quad \|x - \overline{x}\| > \rho - \epsilon \}$$

As $d \to \infty$, the orange collapses to its peel. Hence, no thresholding criterion can discriminate the patterns.

$$\begin{aligned} \mathrm{vol}\left(\mathscr{P}_\epsilon\right) &= \lim_{d \to \infty} \mathrm{vol}\left(\mathscr{N}_\rho\right)\left(1 - \frac{\mathrm{vol}\left(\mathscr{N}_\epsilon\right)}{\mathrm{vol}\left(\mathscr{N}_\rho\right)}\right) \\ &= \mathrm{vol}\left(\mathscr{N}_\rho\right)\left(1 - \lim_{d \to \infty}\left(\frac{\rho - \epsilon}{\rho}\right)^d\right) = \mathrm{vol}\left(\mathscr{N}_\rho\right) \end{aligned}$$

A nice exercise …

Compute the into-char Euclıean distance in MNIST!
You'll lean a lot about space oddities …



Pattern auto-encoding

$$f(w, x_\kappa) \simeq x_\kappa$$

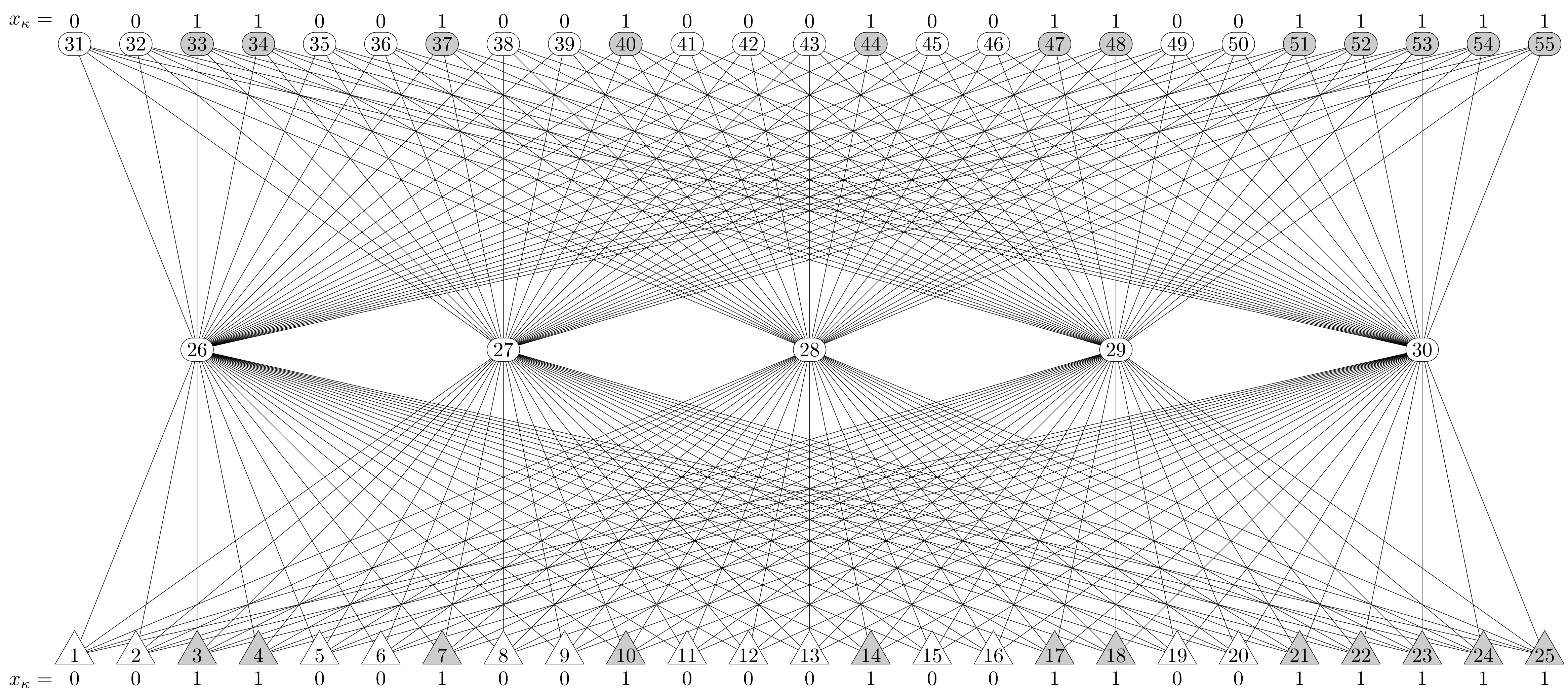


FIGURE 1.4

Pattern auto-encoding by an MLP. The neural net is supervised in such a way to reproduce the input to the output. The hidden layer yields a compressed pattern representation.

Pattern auto-encoding (con't)

$$\mathscr{D} = \{x_1, \ldots, x_\ell\} \subset \mathscr{X}^\ell$$

$$w^\star = \arg\min_w \sum_{x_\kappa \in \mathscr{D}} \|\mathsf{f}(w, x_\kappa) - x_\kappa\|^2$$

$$s_{\mathscr{D}}(x) := \|x - \mathsf{f}(w^\star, x)\|$$

$$\mathscr{X}_{\mathscr{D}}^{\rho} := \{x \in \mathscr{X} \mid s_{\mathscr{D}}(x) < \rho\}$$

Other protocols of learning

- semi-supervised learning
- transductive learning
- reinforcement learning
- active learning