Chapter 2: Supervised Learning

Tyler J. Brough

Data Analytics & Information Systems

Learning a Class from Examples

- Class C of a _"family car"
  - **Prediction:** Is a car x a family car?
  - **Knowledge extraction:** What do people expect from a family car?
- Output:
  - Positive $(+)$ and negative $(-)$ examples
- Input representation:
  - x_1 : price, x_2 : engine power

Training set $\mathcal{X}$

$$\mathcal{X} = \{\mathbf{x}^t, r^t\}_{t=1}^{N}$$

$$r = \begin{cases} 1, & \text{if } \mathbf{x} \text{ is positive} \\ 0, & \text{if } \mathbf{x} \text{ is negative} \end{cases}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Training set $\mathcal{X}$

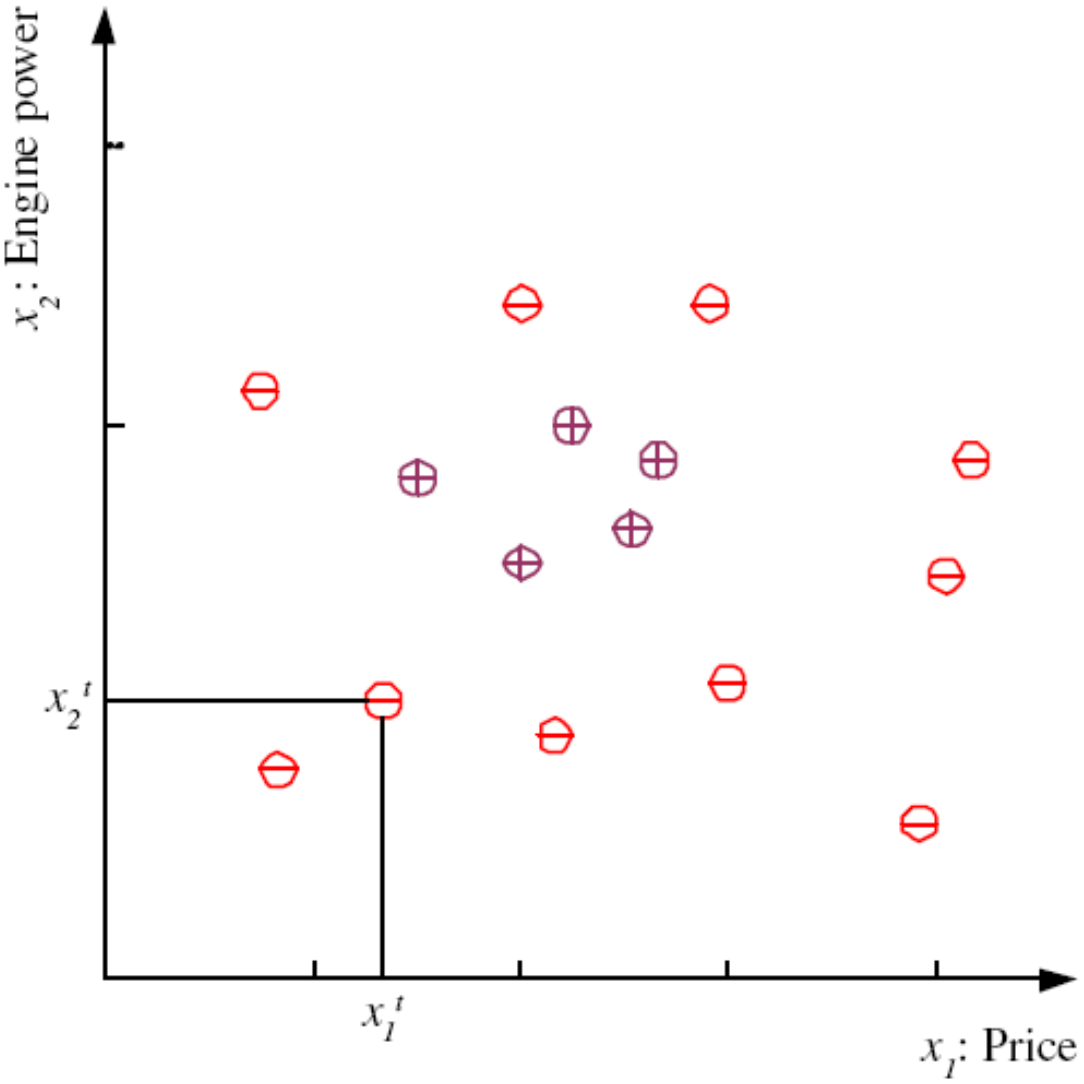


Figure 1: Figure 2.1

Class C

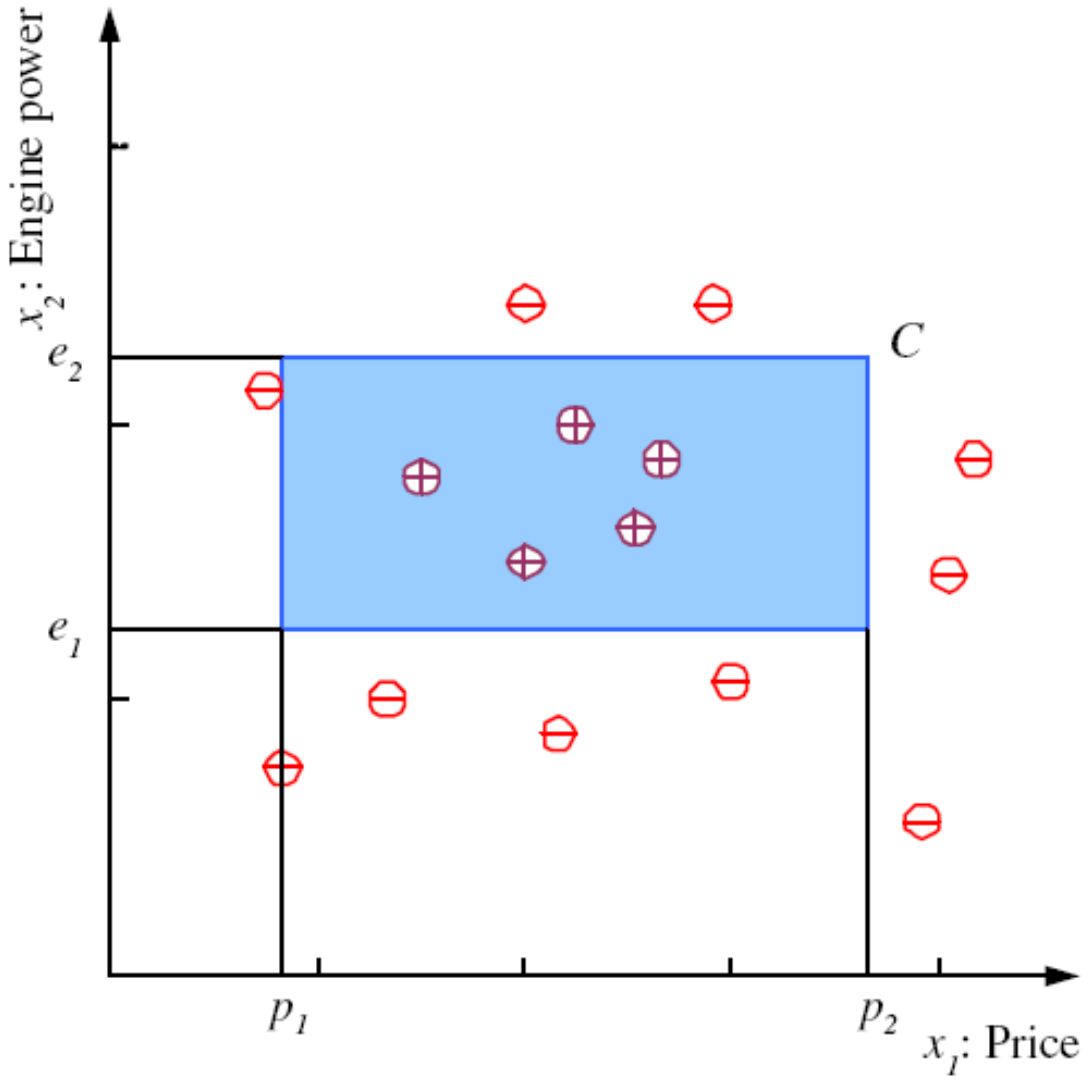


Figure 2: Figure 2.2

Hypothesis class $\mathcal{H}$

$$h(\mathbf{x}) = \begin{cases} 1, & \text{if } h \text{ says } \mathbf{x} \text{ is positive} \\ 0, & \text{if } h \text{ says } \mathbf{x} \text{ is negative} \end{cases}$$

$$E(h|\mathcal{X}) = \sum_{t=1}^{N} 1(h(\mathbf{x}^t) \neq r^t)$$

Hypothesis class $\mathcal{H}$

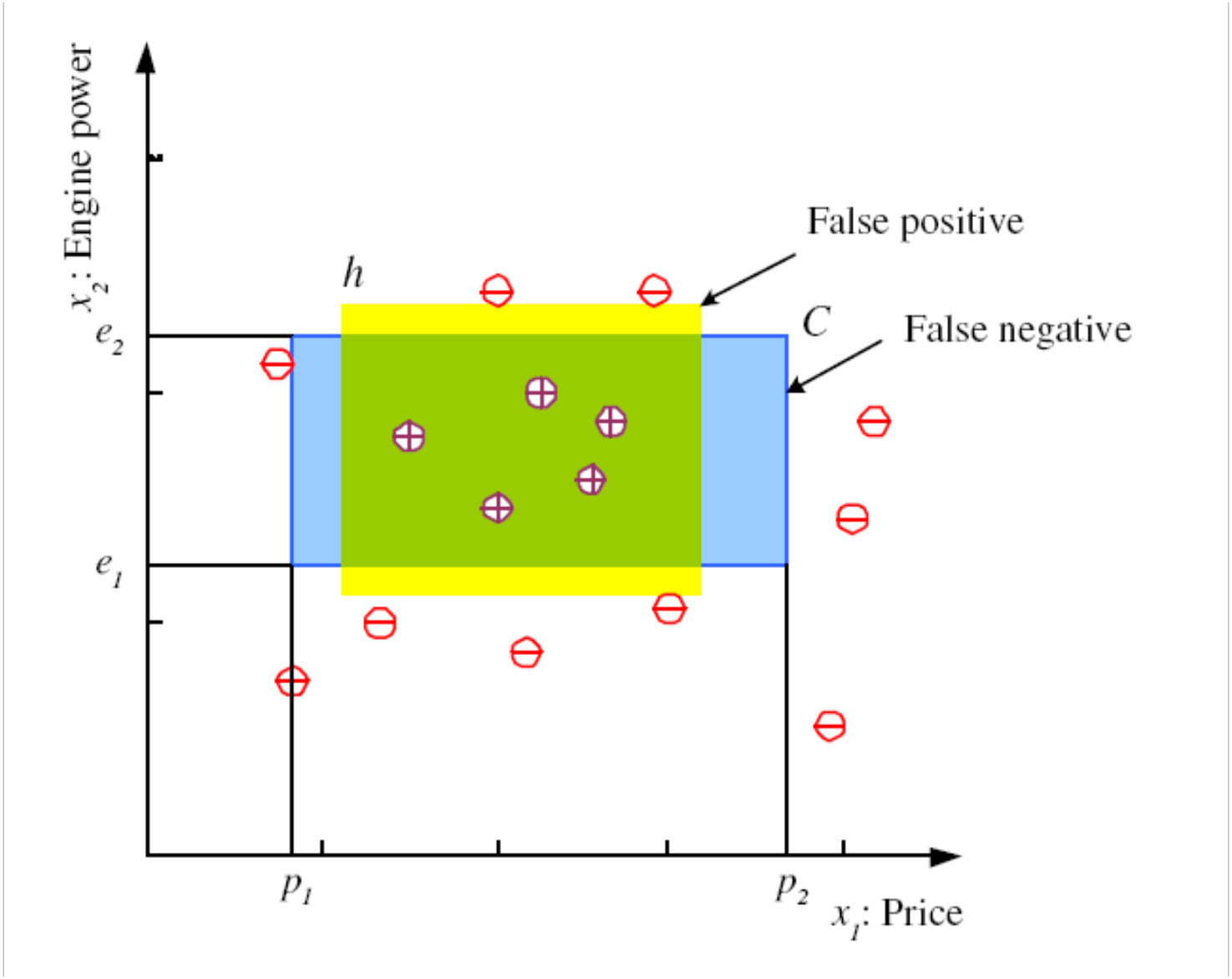


Figure 3: Figure 2.3

S, G, and the Version Space

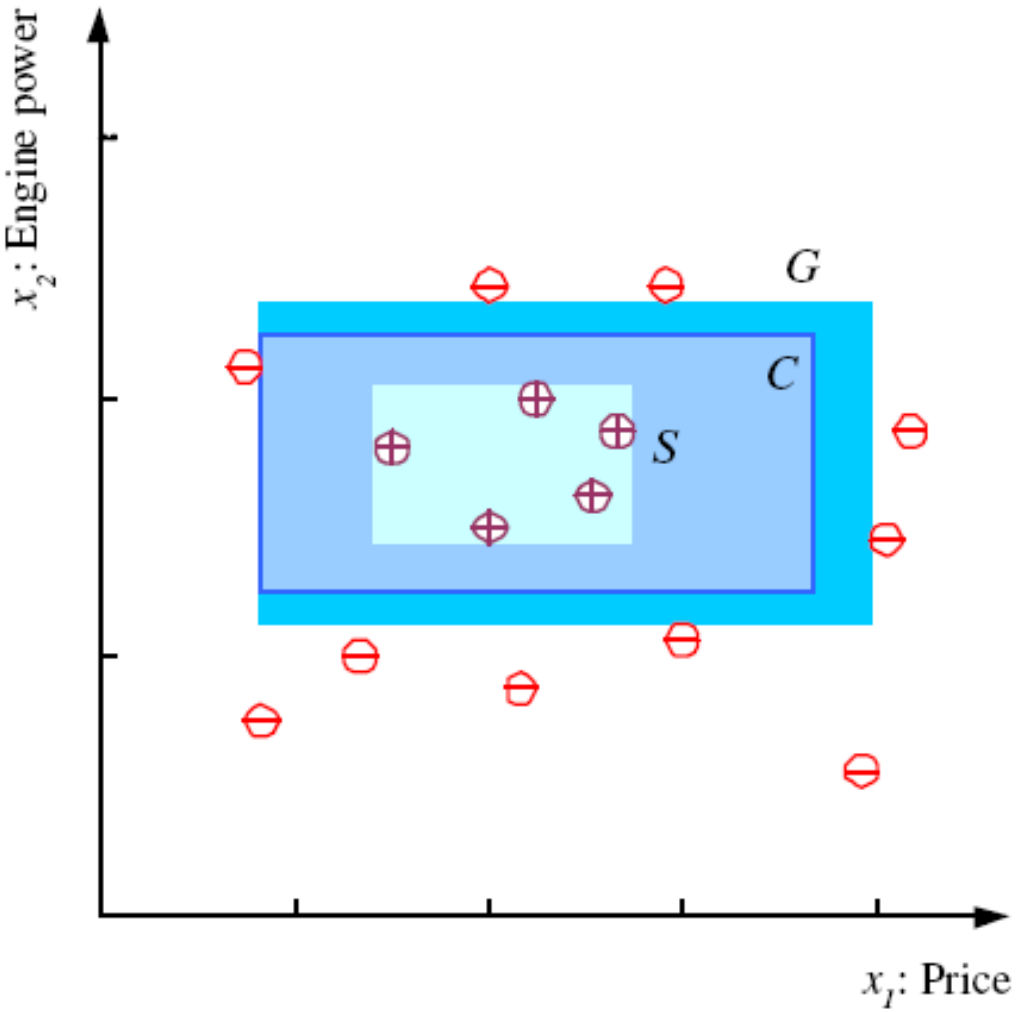


Figure 4: Figure 2.4

Margin

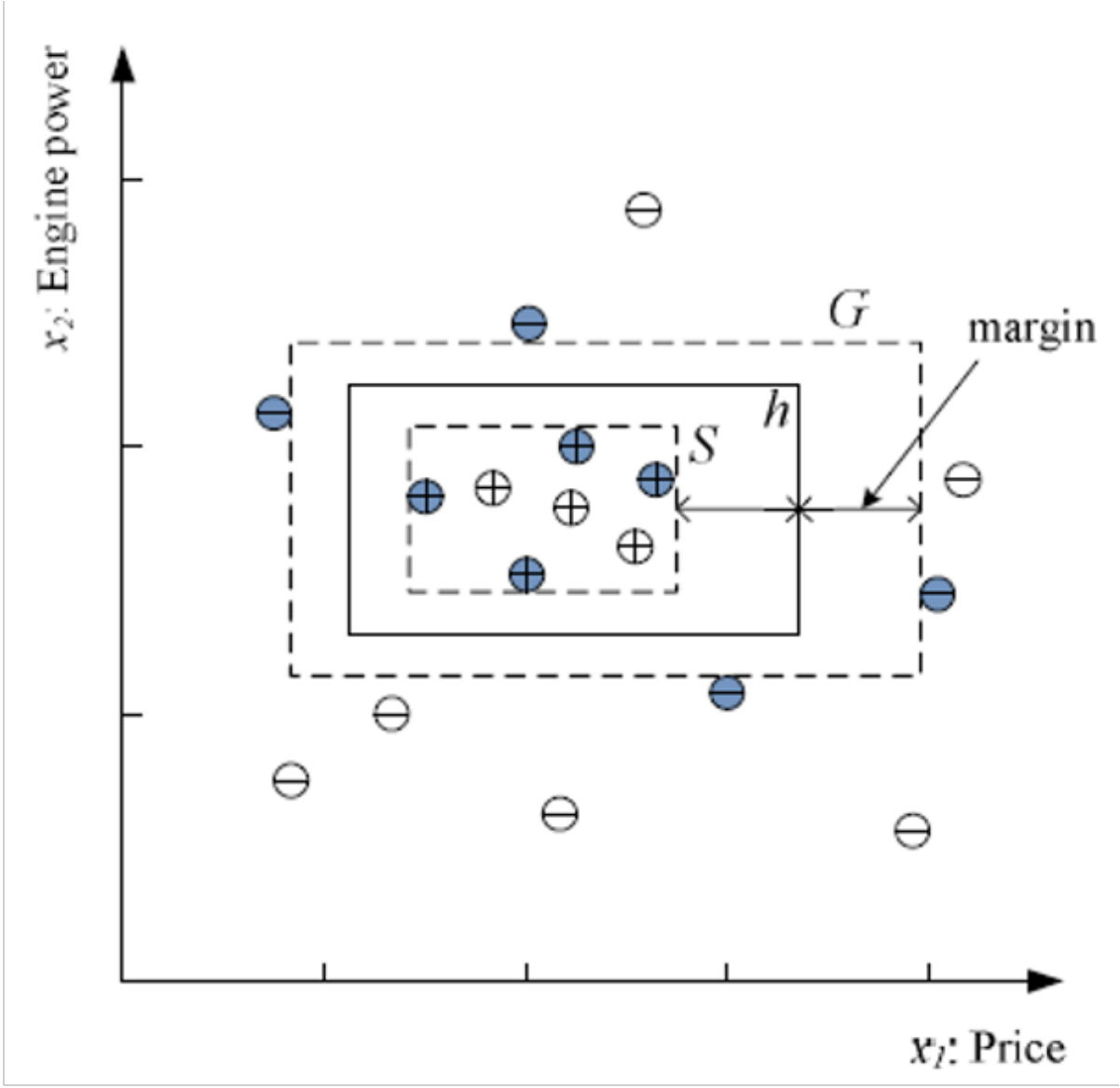
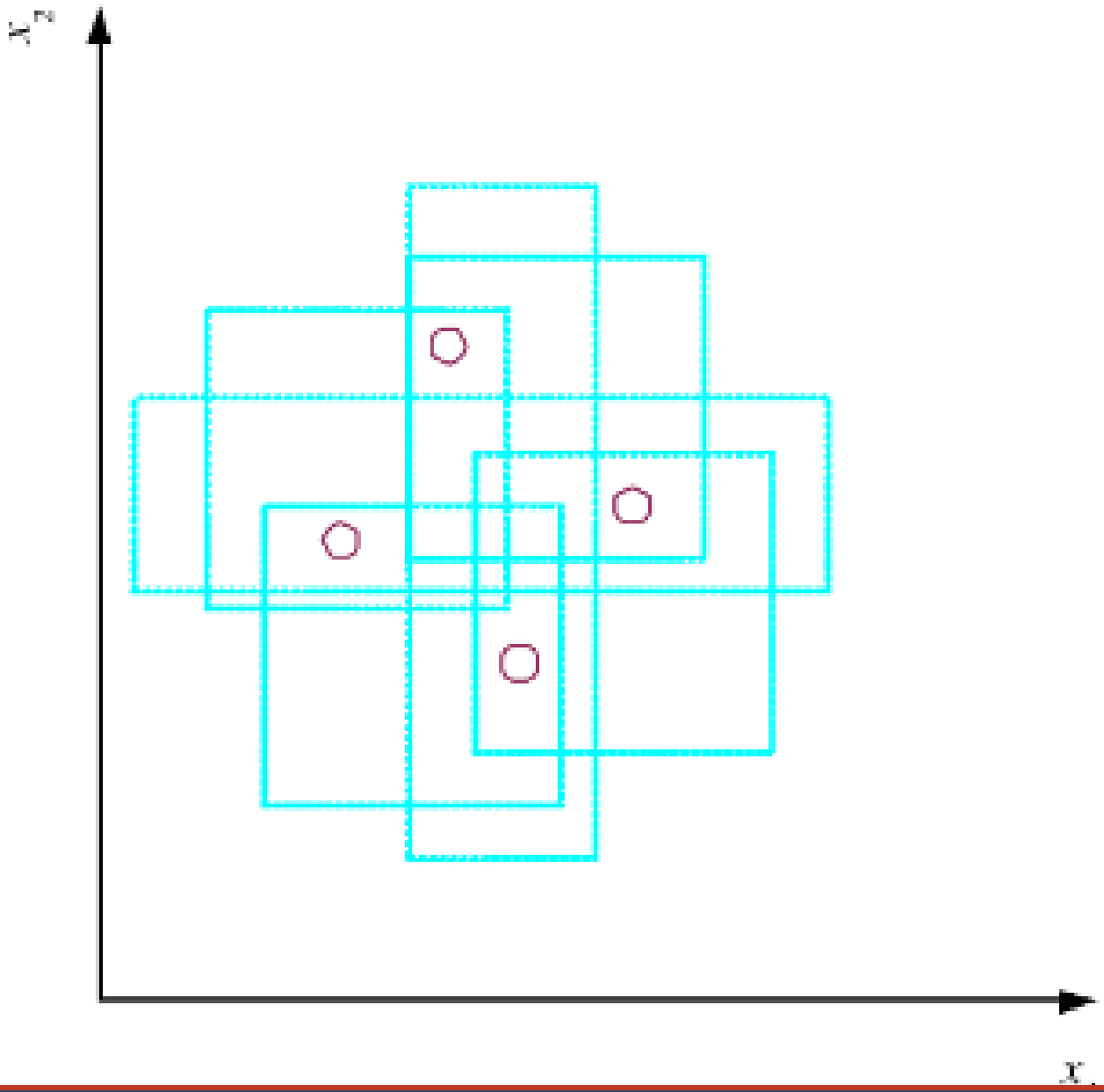


Figure 5: Figure 2.5

VC Dimension

- N points can be labeled in 2^N ways as $+/-$
- $\mathcal{H}$ shatters N if there exists $h \in \mathcal{H}$ consistent for any of these: $VC(\mathcal{H}) = N$

VC Dimention



Probably Approximately Correct (PAC) Learning

- How many training examples N should we have, such that with $1-\delta$, h has **error at most** ϵ? (Blumer et al., 1998)
- Each strip is at most $\epsilon/4$
- Pr that we miss a strip $1-\epsilon/4$
- Pr that N instances miss a strip $(1-\epsilon/4)^N$
- Pr that N instances miss 4 strips $4(1-\epsilon/4)^N$
- $4(1-\epsilon/4)^N \leq \delta$ and $(1-x) \leq \exp(-x)$
- $4\exp(-\epsilon N/4) \leq \delta$ and $N \geq (4/3)\log(4/\delta)$

PAC Learning

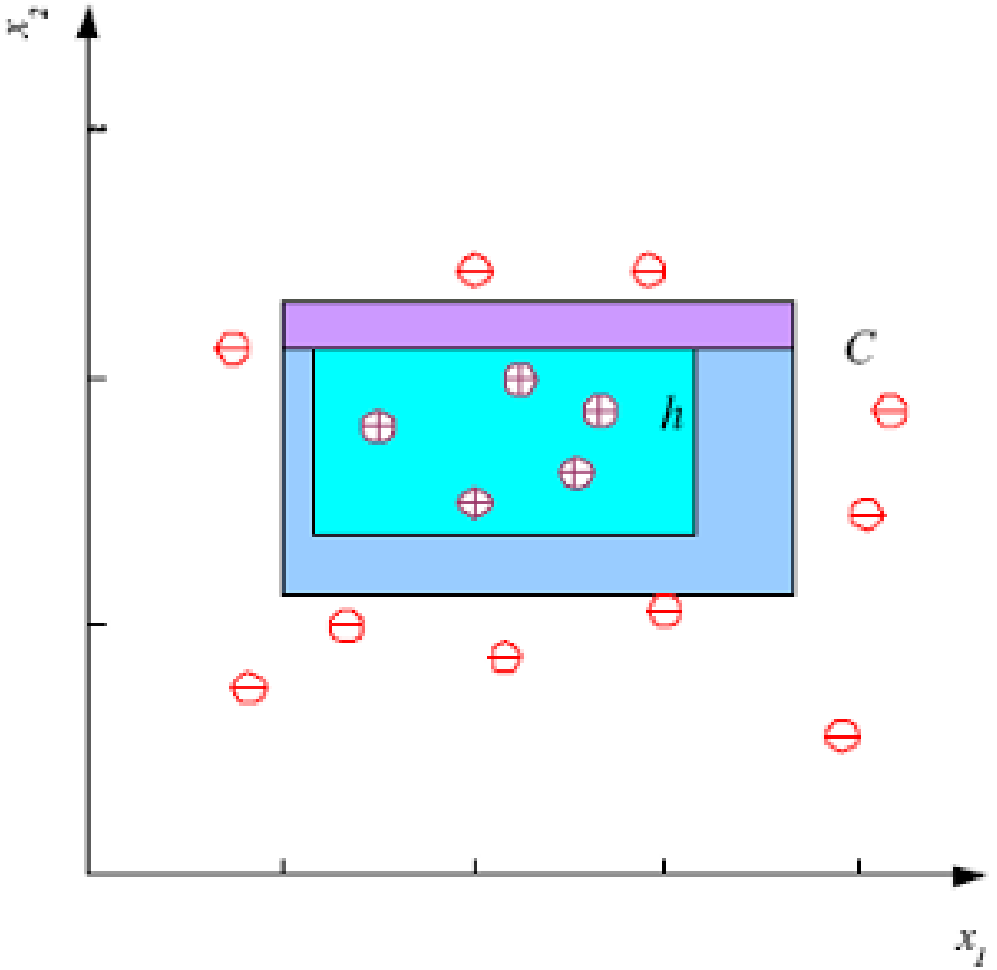


Figure 7: Figure 2.7