

# MACHINE LEARNING

Chao Wang (*Frank*)

**wangchao@nankai.edu.cn**

CHAPTER 2:

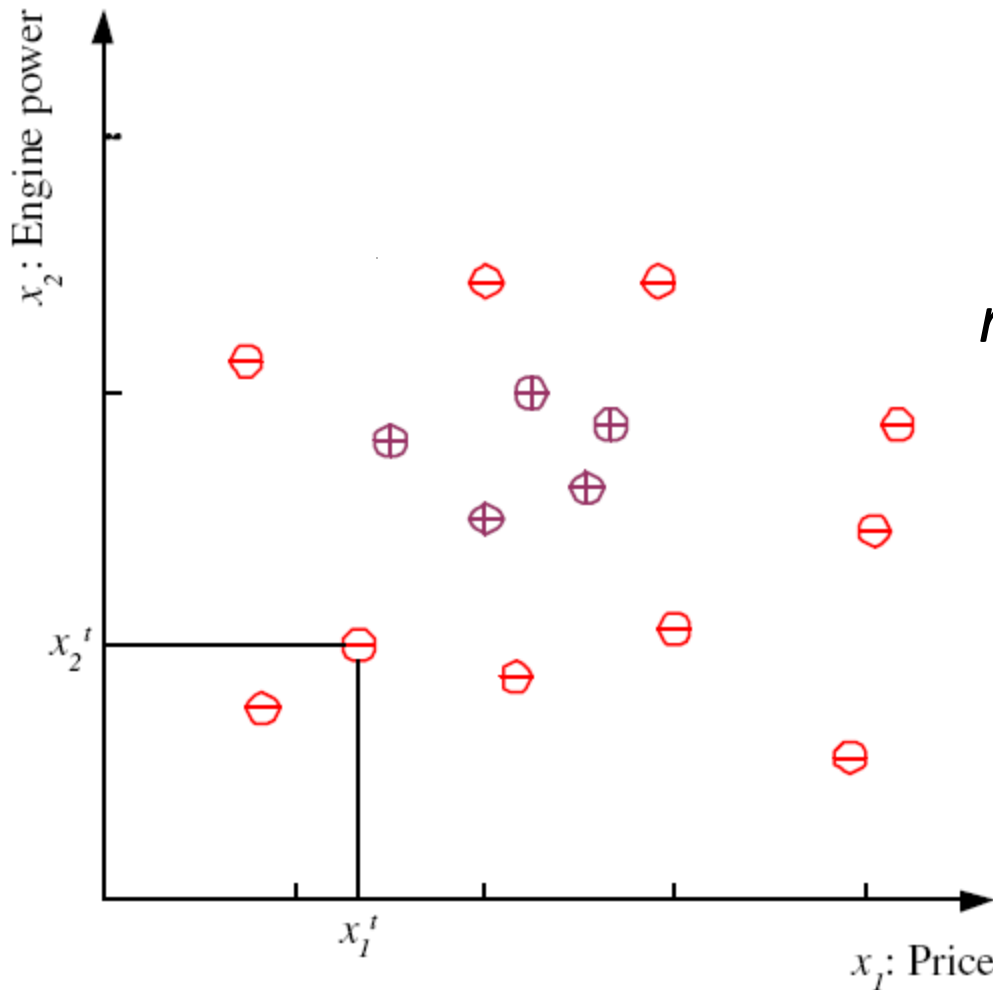
# SUPERVISED LEARNING

# Learning a Class from Examples

3

- Class C of a “family car”
  - ▣ Prediction: Is 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 $\mathcal{X}$



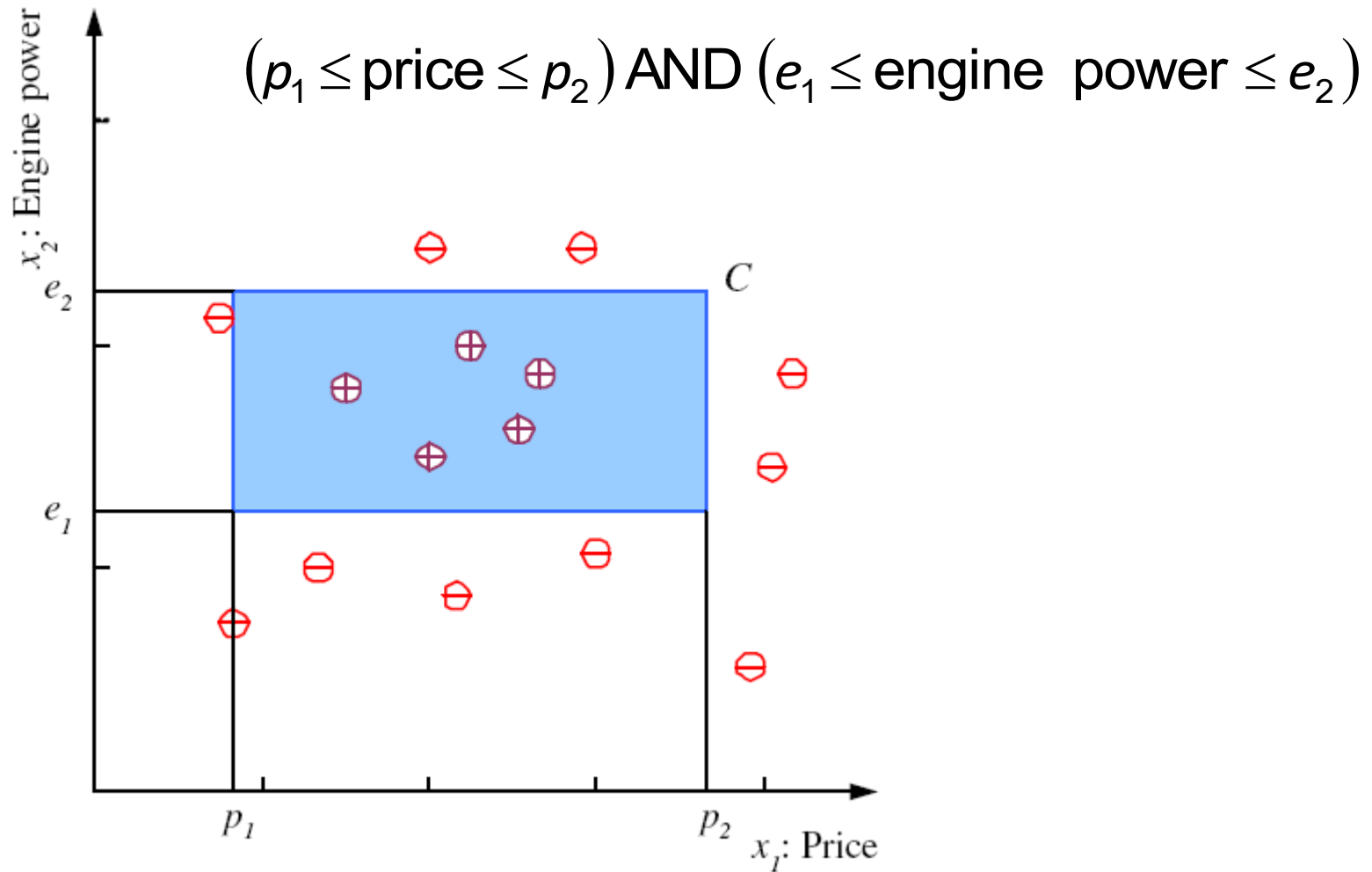
$$\mathcal{X} = \{\mathbf{x}^t, r^t\}_{t=1}^N$$

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

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

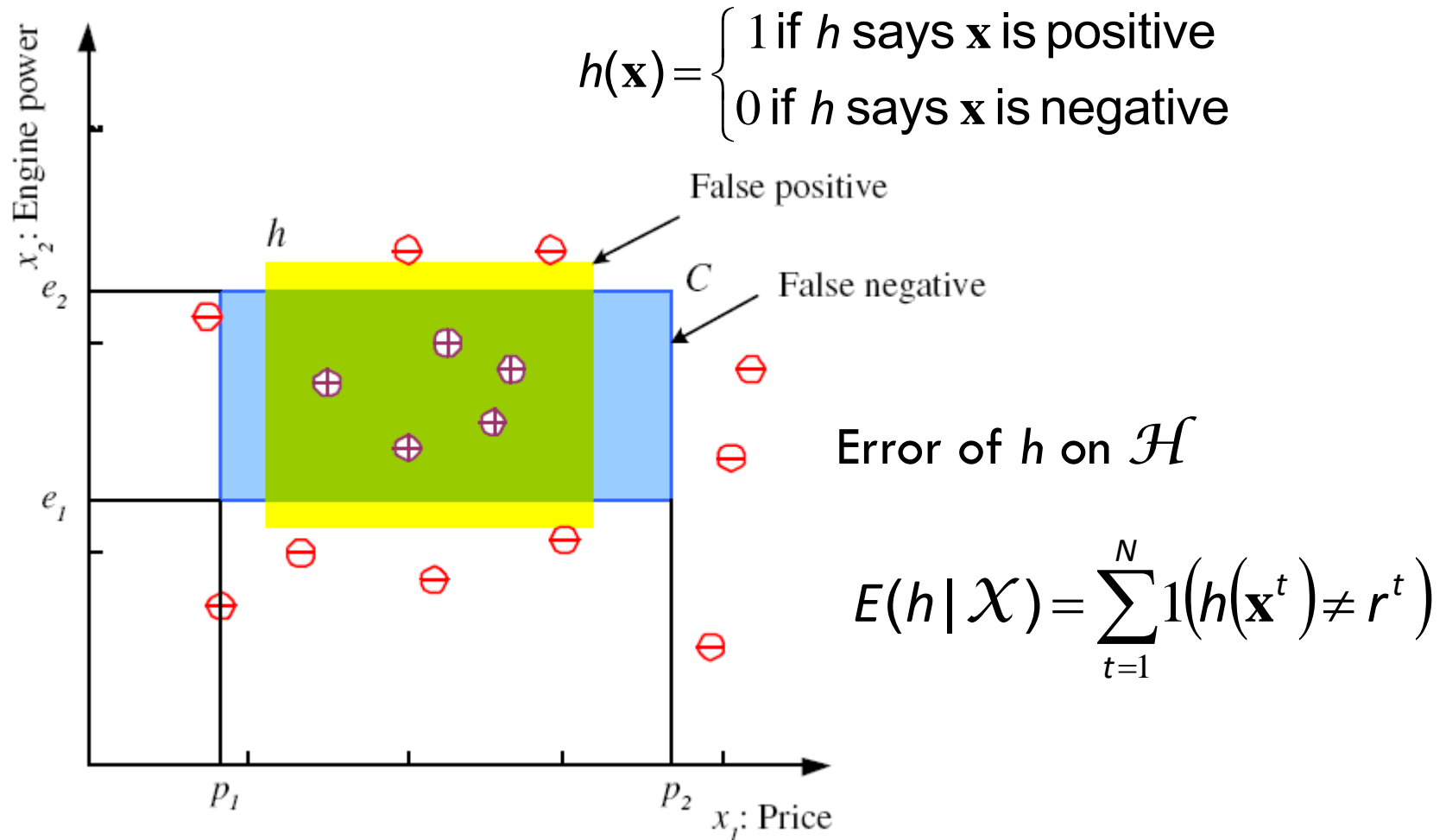
# Class C

5



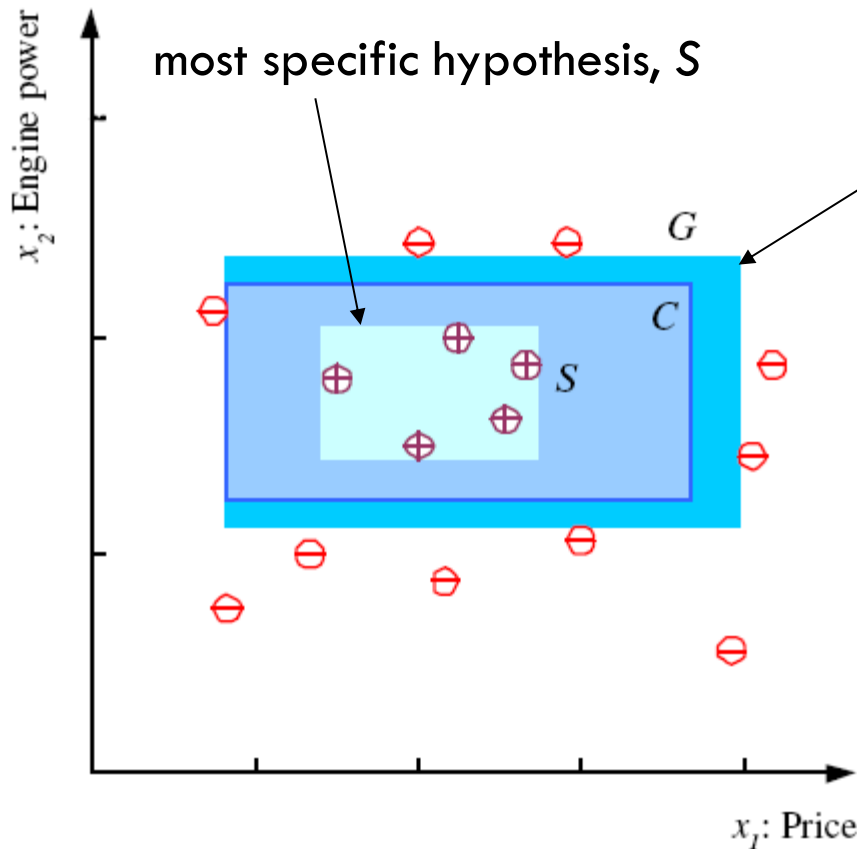
# Hypothesis class $\mathcal{H}$

6



# S, G, and the Version Space

7



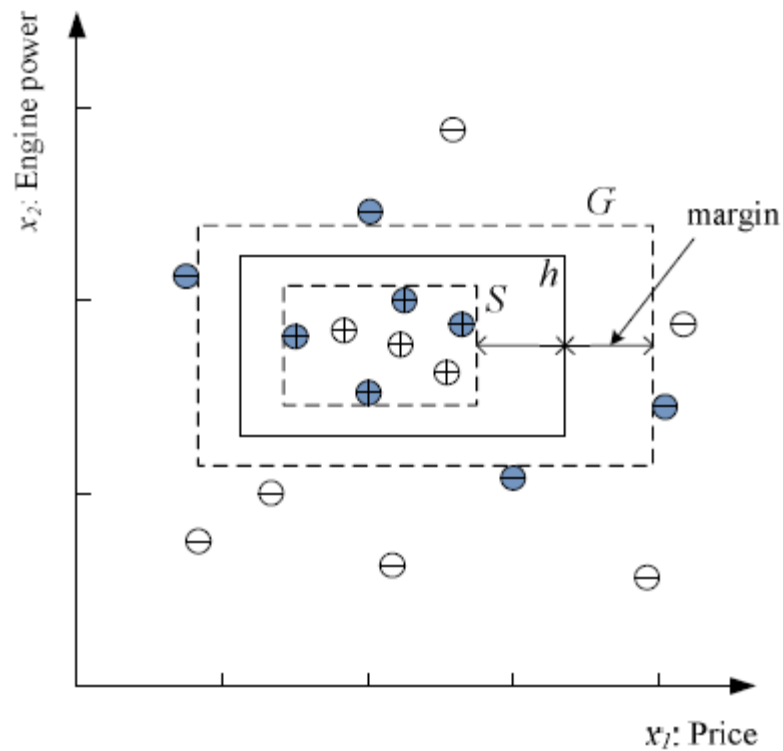
most general hypothesis,  $G$

$h \in H$ , between  $S$  and  $G$  is  
consistent and make up the  
version space  
(Mitchell, 1997)

# Margin

8

- Choose  $h$  with largest margin

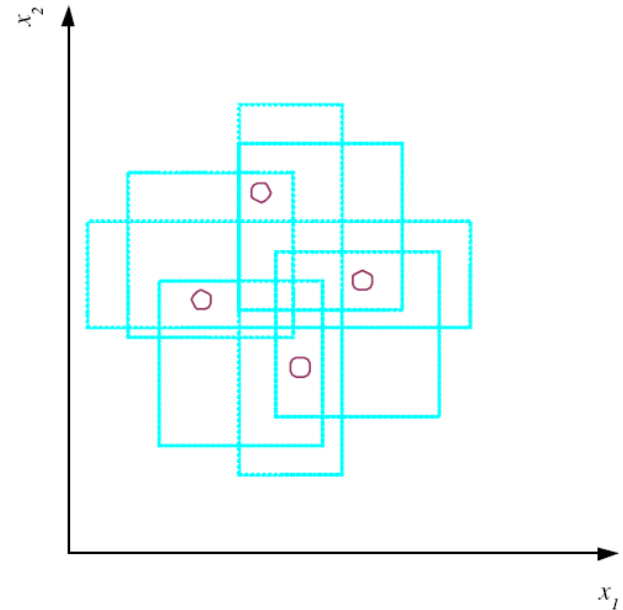




# VC Dimension

9

- $N$  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:  
 $VC(\mathcal{H}) = N$



*An axis-aligned rectangle shatters 4 points only !*

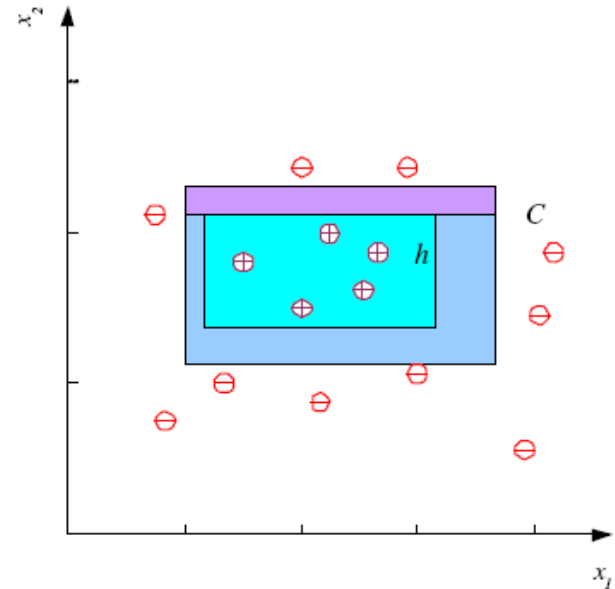
# Probably Approximately Correct (PAC) Learning

10

- How many training examples  $N$  should we have, such that with **probability at least**  $1 - \delta$ ,  $h$  has **error at most**  $\epsilon$  ?

(Blumer et al., 1989)

- 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  $N \geq (4/\epsilon)\log(4/\delta)$

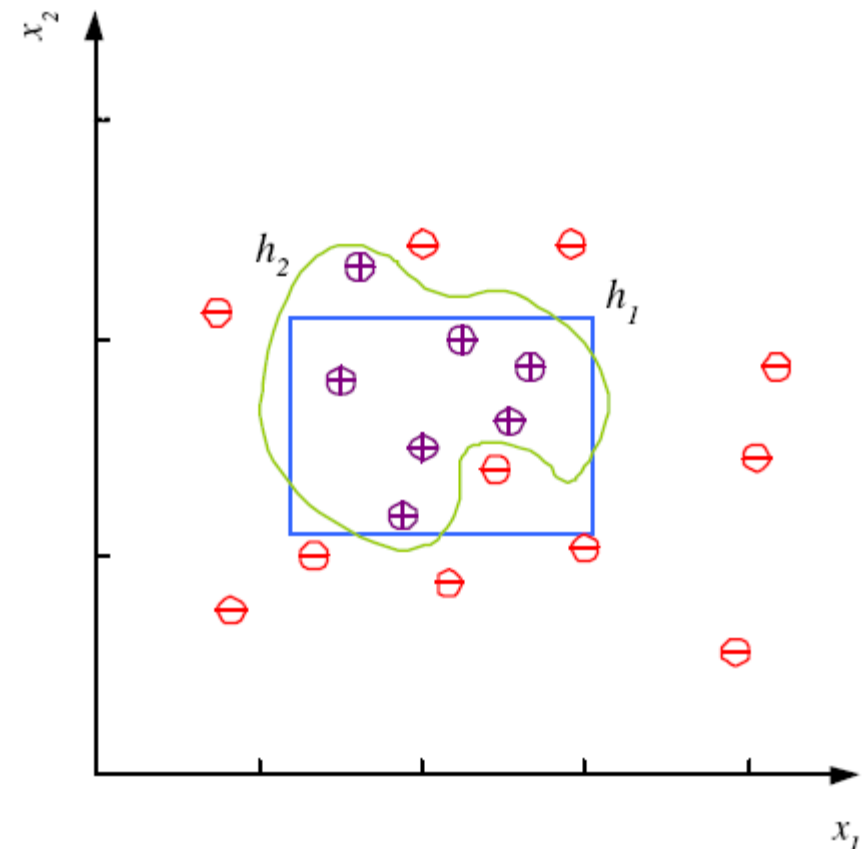


# Noise and Model Complexity

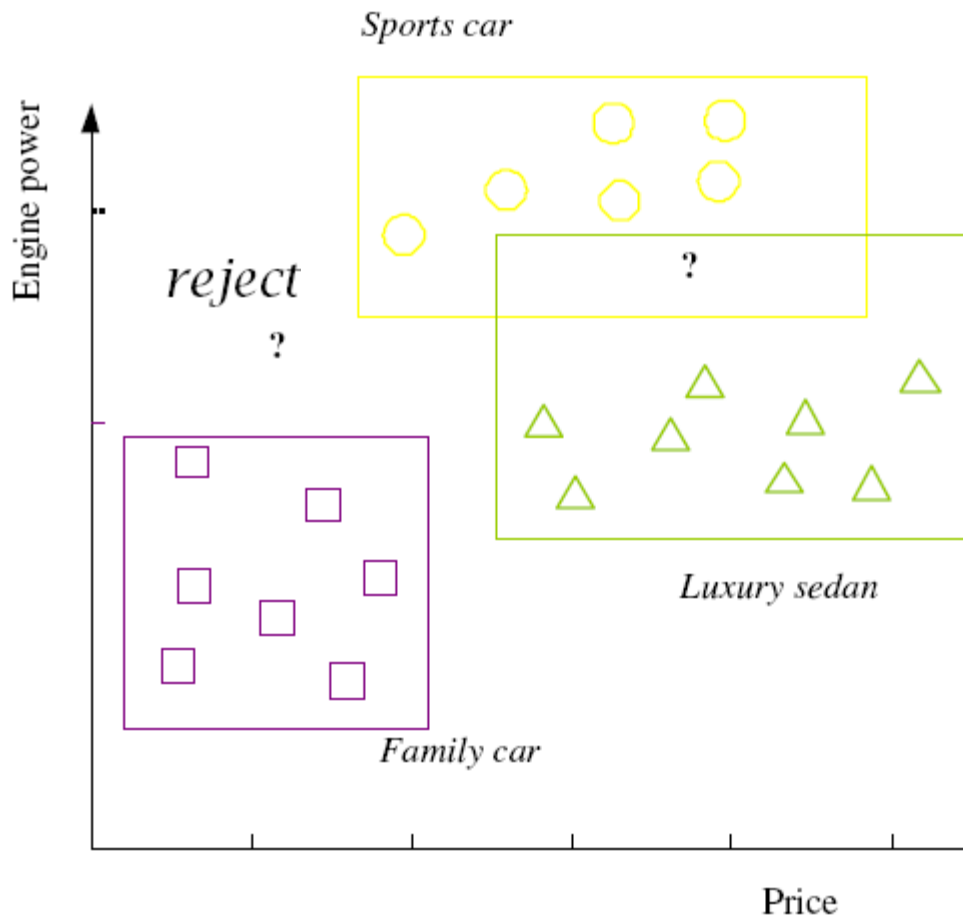
11

Use the simpler one because

- Simpler to use  
(lower computational complexity)
- Easier to train (lower space complexity)
- Easier to explain  
(more interpretable)
- Generalizes better (lower variance - Occam's razor)



# Multiple Classes, $C_i$ $i=1, \dots, K$



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

$$r_i^t = \begin{cases} 1 & \text{if } \mathbf{x}^t \in C_i \\ 0 & \text{if } \mathbf{x}^t \in C_j, j \neq i \end{cases}$$

Train hypotheses

$h_i(\mathbf{x}), i = 1, \dots, K:$

$$h_i(\mathbf{x}^t) = \begin{cases} 1 & \text{if } \mathbf{x}^t \in C_i \\ 0 & \text{if } \mathbf{x}^t \in C_j, j \neq i \end{cases}$$

# Regression

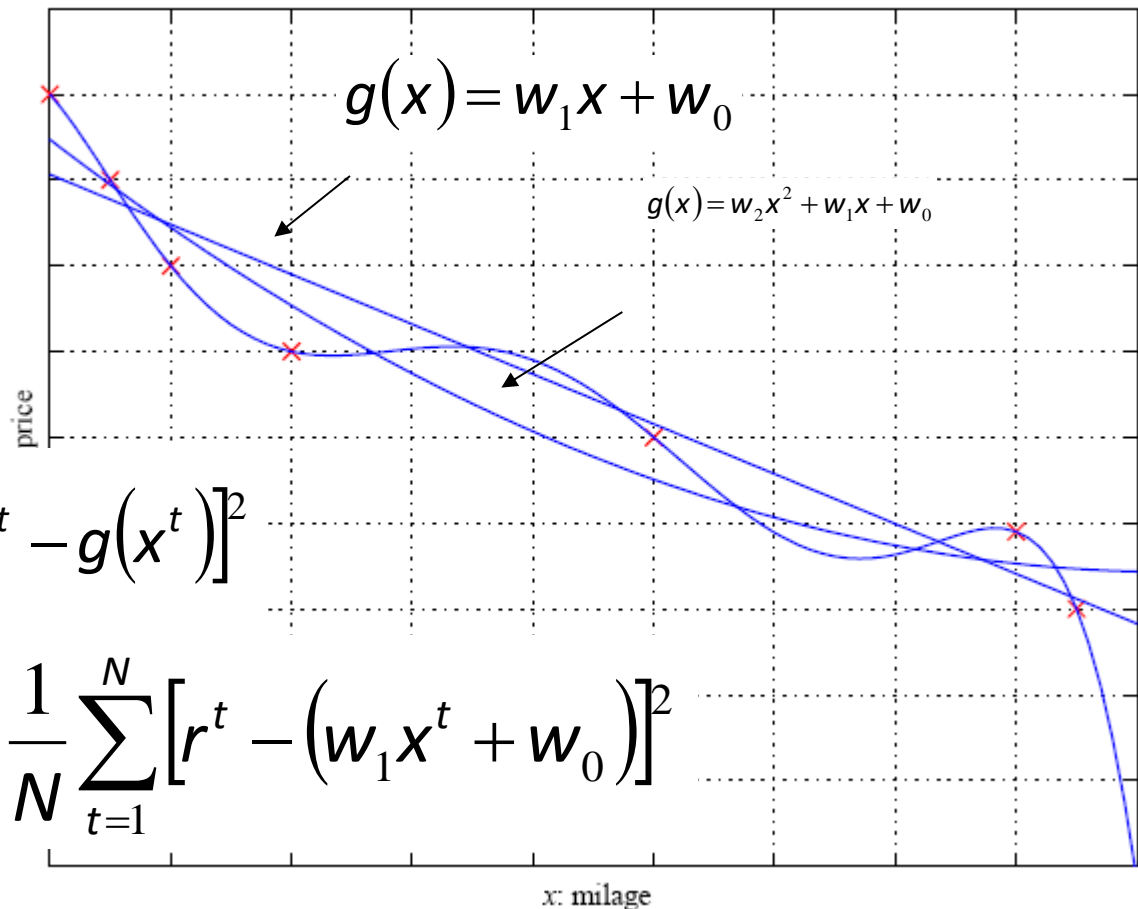
$$\mathcal{X} = \{x^t, r^t\}_{t=1}^N$$

$$r^t \in \mathbb{R}$$

$$r^t = f(x^t) + \varepsilon$$

$$E(g | \mathcal{X}) = \frac{1}{N} \sum_{t=1}^N [r^t - g(x^t)]^2$$

$$E(w_1, w_0 | \mathcal{X}) = \frac{1}{N} \sum_{t=1}^N [r^t - (w_1 x^t + w_0)]^2$$



# Model Selection & Generalization

14

- Learning is an ill-posed problem; if data is not sufficient to find a unique solution
- The need for inductive bias, assumptions about  $\mathcal{H}$
- Generalization: How well a model performs on new data
- Overfitting:  $\mathcal{H}$  more complex than  $C$  or  $f$
- Underfitting:  $\mathcal{H}$  less complex than  $C$  or  $f$

# Triple Trade-Off

15

- There is a trade-off between three factors (Dietterich, 2003):
  1. Complexity of  $\mathcal{H}$ ,  $c(\mathcal{H})$ ,
  2. Training set size,  $N$ ,
  3. Generalization error,  $E$ , on new data
- As  $N \uparrow$ ,  $E \downarrow$
- As  $c(\mathcal{H}) \uparrow$ , first  $E \downarrow$  and then  $E \uparrow$

# Cross-Validation

16

- To estimate generalization error, we need data unseen during training. We split the data as
  - ▣ Training set (50%)
  - ▣ Validation set (25%)
  - ▣ Test (publication) set (25%)
- Resampling when there is few data



# Dimensions of a Supervised Learner

1. Model:  $g(\mathbf{x} | \theta)$

2. Loss function:  $E(\theta | \mathcal{X}) = \sum_t L(r^t, g(\mathbf{x}^t | \theta))$

3. Optimization procedure:

$$\theta^* = \arg \min_{\theta} E(\theta | \mathcal{X})$$

*Questions?*