Supervised Learning: $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$

x_i → ith data point

y_i → ith label

☆ predicting whether a car is a family car or not

$\underbrace{[\;]}$ ⇒ price, engine power

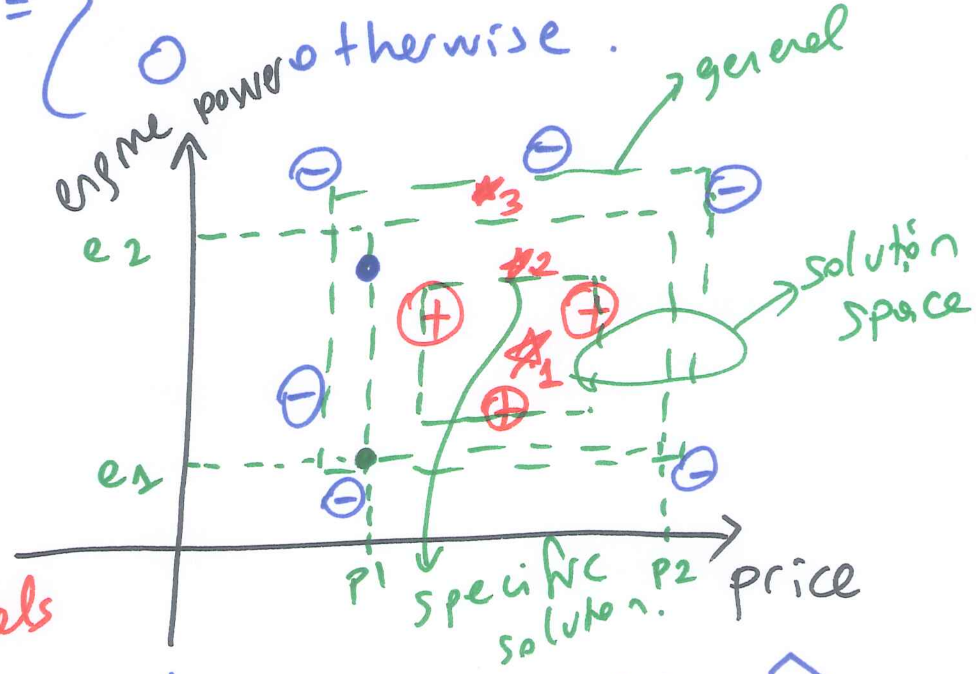
price, engine power

$$\mathcal{X} \Rightarrow \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_9 \end{array} \begin{bmatrix} \quad \\ \quad \\ \quad \\ \quad \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} \begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_9 \end{array}$$

9×2 ⇒ data matrix ($\mathbf{X}$)

9×1 ⇒ label vector ($\mathbf{y}$)

$$y_i = \begin{cases} 1 & \text{if } x_i \text{ is a family car} \\ 0 & \text{otherwise.} \end{cases}$$



⊕ vs ⊖ using rectangles

family of models

the set of parameters

$\theta = \{p_1, p_2, e_1, e_2\}$

$f(x_{N+1} \mid p_1, p_2, e_1, e_2) = \hat{y}_{10}$

$$f(x_{N+1} \mid p_1, p_2, e_1, e_2) = \begin{cases} 1 & p_1 < x_{N+1,1} < p_2 \;\&\; e_1 < x_{N+1,2} < e_2 \\ 0 & \text{otherwise} \end{cases}$$

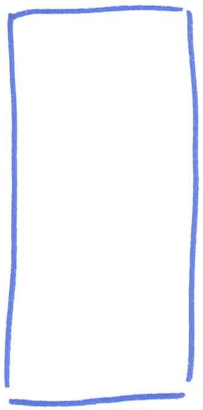
★1 ★2	TRUE	TRUE
★3	TRUE	FALSE

$$f(x_{N+1} \mid \Theta)$$

TRAINING: learning (deciding) $p_1, p_2, e_1, \& e_2$

TESTING: checking whether a data point is in the inferred rectangle using $\underline{\underline{\Theta}}$

thin and tall



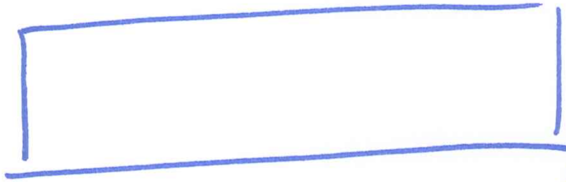
X_1

$N \times D$

$N \gg D$

$|\Theta| = \underline{2 \times D}$

wide and short

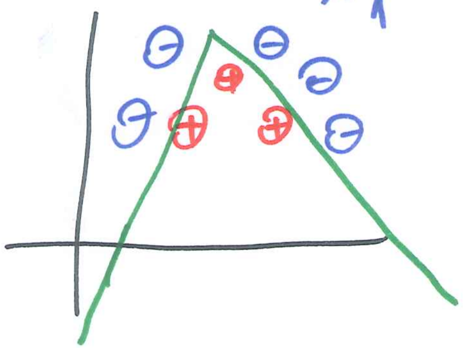


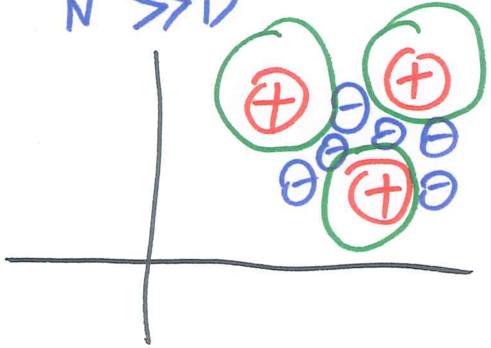
X_2

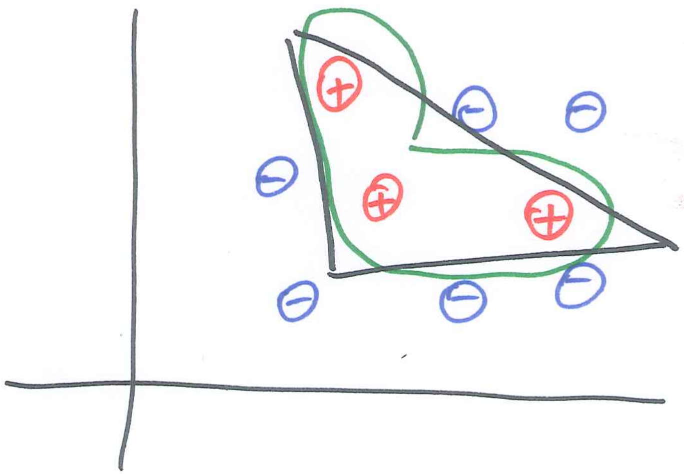
$\underline{N} \times D$

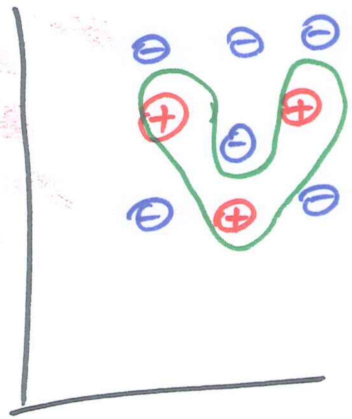
$D \gg N$

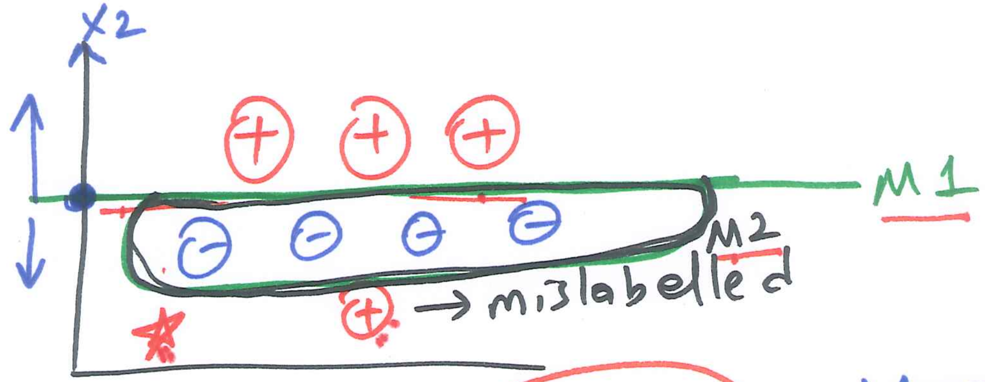
{ ★ model complexity
{ ★ prediction performance











☆ is $\ominus$ M1

☆ is $\oplus$ M2

{ training data

{ test data

TRADE-OFF

objective function = MODEL COMPLEXITY + GENERALIZATION ERROR

minimize.

1/8

0/8

{ ↑ high

{ ↓ low

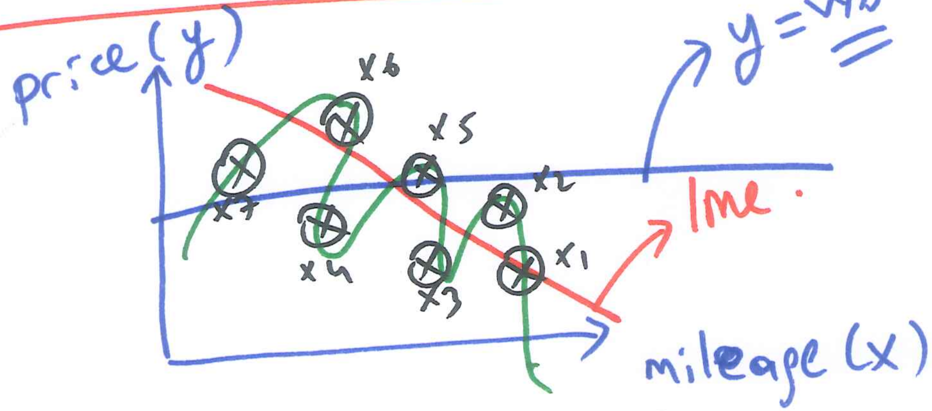
training.

testing

training accuracy of M1 : 7/8

M2 : 8/8

Linear Regression:



Slope, intercept

$$x = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{7\times 1}$$

mileage

data matrix

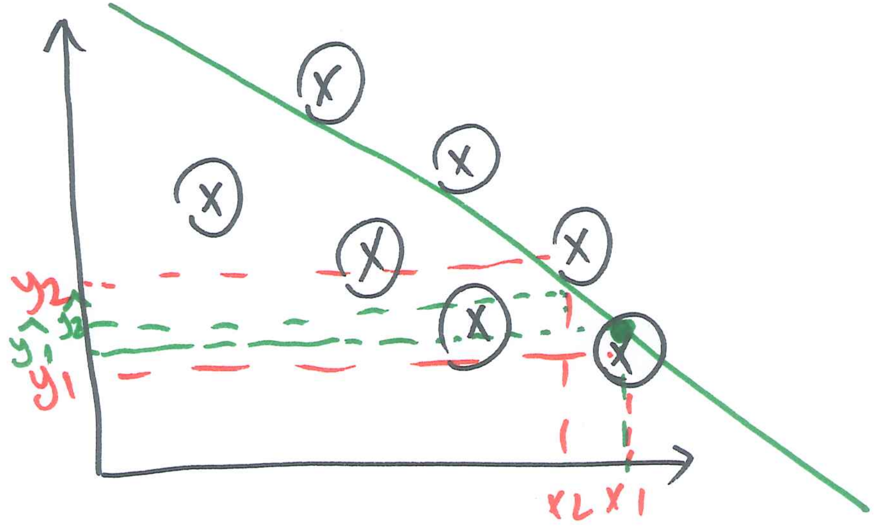
$$y = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{7\times 1}$$

price.

labels.

$$\hat{y} = \underbrace{a}_{\text{slope}} x + \underbrace{b}_{\text{intercept}}$$

$$\hat{y} = \underbrace{w}_{\text{weight}} x + w_0$$

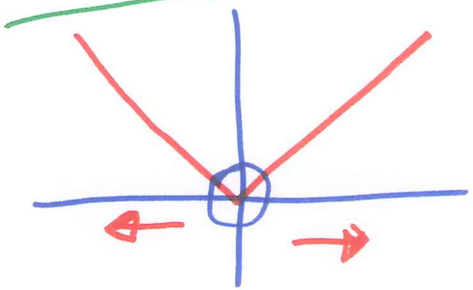


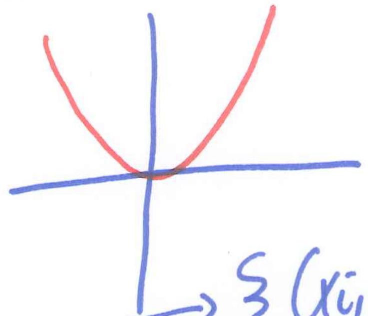
x_1	y_1	$\hat{y}_1$	$\hat{y}_1 - y_1$
x_2	y_2	$\hat{y}_2$	$y_2 - \hat{y}_2$
$\vdots$	$\vdots$	$\vdots$	
x_7	y_7	$\hat{y}_7$	
	observed values	predicted values.	

~~minimize~~ $\cancel{\sum_{i=1}^{N} (y_i - \hat{y}_i)}$

? minimize $\sum_{i=1}^{N} |y_i - \hat{y}_i|$

minimize $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$





minimize $\sum_{i=1}^{N} (y_i - w x_i - w_0)^2 = \text{ERROR}(w, w_0 \mid \mathcal{X})$

$\mathcal{X} \rightarrow \{(x_i, y_i)\}_{i=1}^{N}$

$$\text{Error}(w, w_0 \mid \mathcal{X}) = \sum_{i=1}^{N} (y_i - w x_i - w_0)^2 \rightarrow \text{squared error}$$

$$\frac{\partial \text{Error}}{\partial w} = \frac{\partial \sum_{i=1}^{N} (y_i - w x_i - w_0)^2}{\partial w} = \sum_{i=1}^{N} \frac{\partial (y_i - w x_i - w_0)^2}{\partial w}$$

$$0 = \sum_{i=1}^{N} 2.(y_i - w x_i - w_0) . -x_i$$

$$\frac{\partial \text{Error}}{\partial w_0} = \frac{\partial \sum_{i=1}^{N} (y_i - w x_i - w_0)^2}{\partial w_0} = \sum_{i=1}^{N} 2.(y_i - w x_i - w_0)(-1) = 0$$

Exercise #1: Solve these equations for w & w_0.

$$w = \frac{\sum_{i=1}^{N} x_i y_i - \left(\overbrace{\frac{\sum_{i=1}^{N} x_i}{N}}^{\bar{x}}\right) . \left(\overbrace{\frac{\sum_{i=1}^{N} y_i}{N}}^{\bar{y}}\right) . N}{\sum_{i=1}^{N} x_i^2 - N \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2}$$

$$w_0 = \left(\frac{\sum_{i=1}^{N} y_i}{N}\right) - w . \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)$$

ML Algorithm

0) collect data set
1) pick a model family
2) pick a loss function
3) learn parameters.