Supervised Learning

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

x_i → i^{th} data point

y_i → i^{th} label, i^{th} output

* predicting whether a car is a family car or not

car → x_i ; family car or not → y_i

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \begin{matrix} \to \text{price} \\ \to \text{engine power} \end{matrix}$$

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

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{bmatrix} \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{matrix}$$

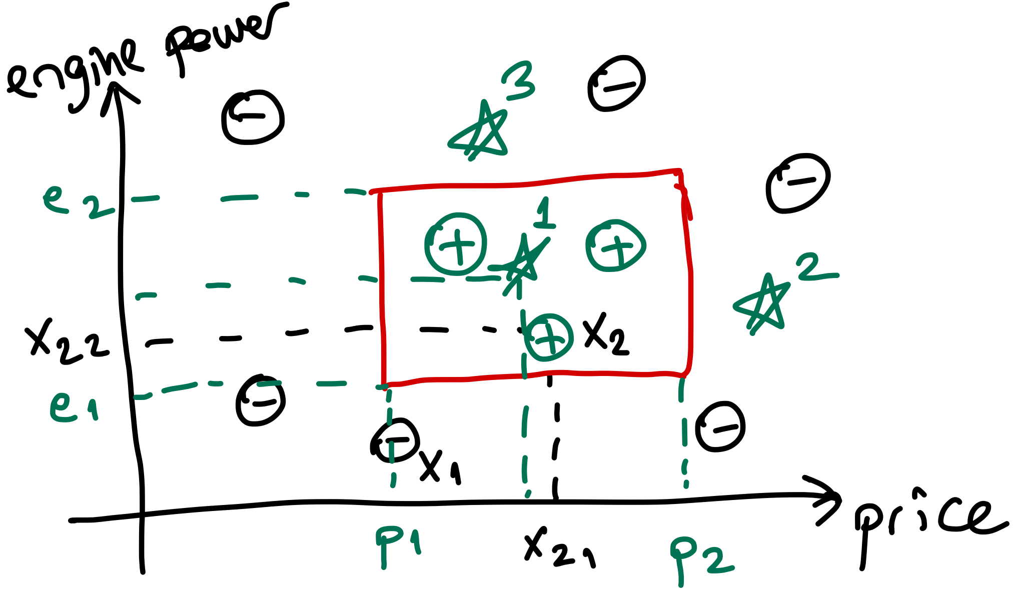
data matrix

$N \times D$ → D: # of features, N: # of data points

in our case $D = 2$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

label vector



9 cars → 3 family cars
→ 6 other types of cars

$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \quad y_1 = 0$$

$$x_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \quad y_2 = 1$$

$$\underbrace{\text{RECTANGLES}}_{\text{model family}}$$

$$\underbrace{\theta}_{\text{model parameters}} = \{p_1, p_2, e_1, e_2\}$$

Learning: finding θ^*

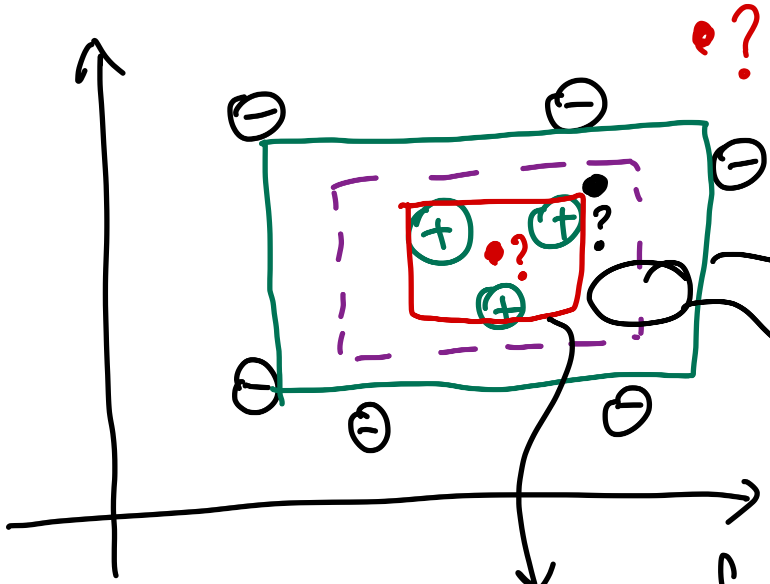
$$f(x_{N+1} \mid p_1, p_2, e_1, e_2) = ? \Leftarrow \text{prediction}$$

$$= \begin{cases} 1 & \text{if } p_1 \leq x_{N+1,1} \leq p_2 \\ & \& \ e_1 \leq x_{N+1,2} \leq e_2 \\ 0 & \text{otherwise} \end{cases}$$

$\star^1$: TRUE & TRUE $\Rightarrow$ 1

$\star^2$: FALSE & TRUE $\Rightarrow$ 0

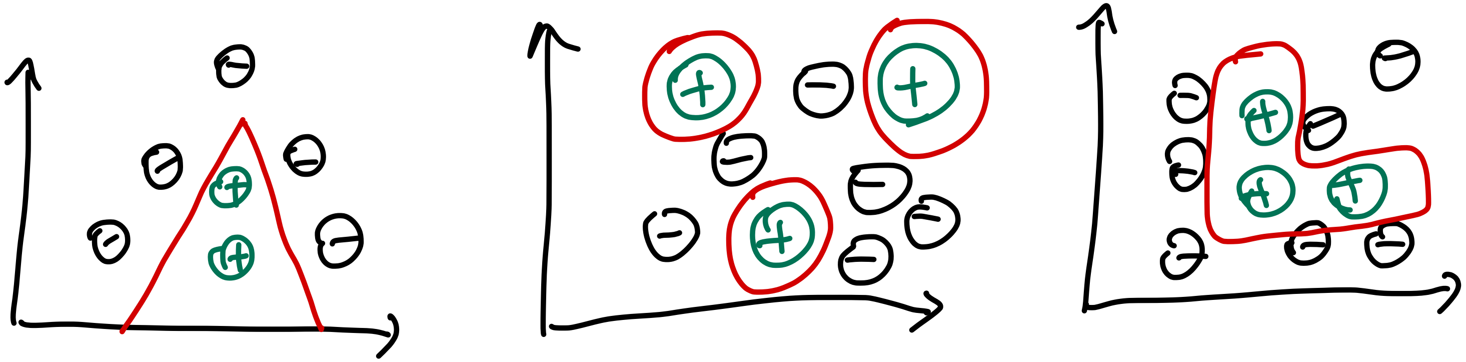
$\star^3$: TRUE & FALSE $\Rightarrow$ 0



general solution

all rectangles in between are called "version space"

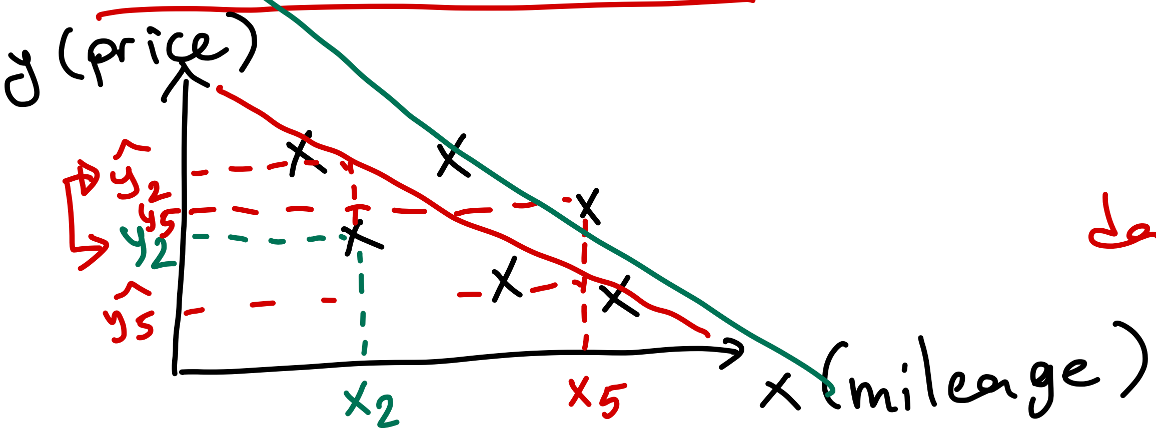
specific solution



MODEL COMPLEXITY ↑

PREDICTION PERFORMANCE (TRAINING ↑ / TEST ↓)

Linear Regression:



$$X = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{61} \end{bmatrix}_{6 \times 1} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{bmatrix}_{6 \times 1}$$

data matrix

6×1: (N) (D)

output vector

$$\theta = \{w_0, w_1\}$$

Set of lines

$\underbrace{\text{Set of lines}}_{\text{MODEL FAMILY.}}$

$$\hat{y} = w_0 + w_1 x$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_6 \end{bmatrix} \begin{matrix} = w_0 + w_1 . x_1 \\ = w_0 + w_1 . x_2 \\ \vdots \\ = w_0 + w_1 . x_6 \end{matrix}$$

predictions

$$\boxed{\begin{matrix} e_1 \\ e_2 \\ \vdots \\ e_6 \end{matrix}} \begin{matrix} = y_1 - \hat{y}_1 \\ = y_2 - \hat{y}_2 \\ \vdots \\ = y_6 - \hat{y}_6 \end{matrix}$$

errors

OPT 1

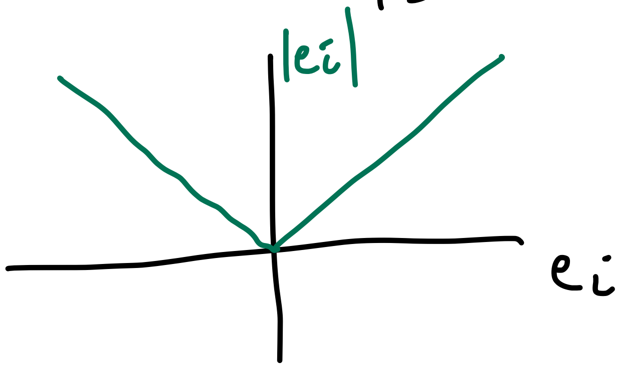
minimize $\sum_{i=1}^{N}(y_i-\hat{y}_i)=\sum_{i=1}^{N} e_i$

×

OPT 2

minimize $\sum_{i=1}^{N}|y_i-\hat{y}_i| = \sum_{i=1}^{N}|e_i| = \sum_{i=1}^{N}|y_i - w_0 - w_1 x_i|$

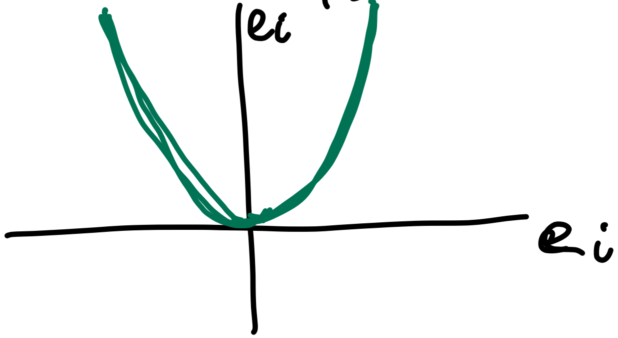
×



OPT 3

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

✓



minimize $\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{N}(y_i-w_0-w_1x_i)^2$

with respect to: w_0, w_1

$$\text{Error}(w_0,w_1 \mid \mathcal{X}) = \sum_{i=1}^{N}(y_i-w_0-w_1x_i)^2$$

$$\frac{\partial \text{Error}}{\partial w_0} = \frac{\partial \sum_{i=1}^{N}(y_i-w_0-w_1x_i)^2}{\partial w_0} = \sum_{i=1}^{N}\frac{\partial (y_i-w_0-w_1x_i)^2}{\partial w_0}$$

$$= \sum_{i=1}^{N} 2.(-1).(y_i-w_0-w_1x_i)$$

$$= \sum_{i=1}^{N} 2(w_0+w_1x_i-y_i)$$

$$\frac{\partial \text{Error}}{\partial w_1} = \sum_{i=1}^{N} 2x_i(w_0+w_1x_i-y_i)$$

EXERCISE #1: Solve for w_0 and w_1.

$$w_1 = \frac{\sum_{i=1}^{N} x_i y_i - \underbrace{\left(\sum_{i=1}^{N} x_i / N\right)}_{\bar{x}} \underbrace{\left(\sum_{i=1}^{N} y_i / N\right)}_{\bar{y}} N}{\sum_{i=1}^{N} x_i^2 - N \Big(\underbrace{\sum_{i=1}^{N} x_i / N}_{\bar{x}}\Big)^2}$$

$$w_0 = \underbrace{\left(\sum_{i=1}^{N} y_i / N\right)}_{\bar{y}} - w_1 \cdot \underbrace{\left(\sum_{i=1}^{N} x_i / N\right)}_{\bar{x}} = \bar{y} - w_1 . \bar{x}$$

ML ALGORITHM

1. collect data $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$
2. pick a model family $\Rightarrow$ set of lines
3. pick a loss/error function $\Rightarrow$ squared error
4. learn the parameters