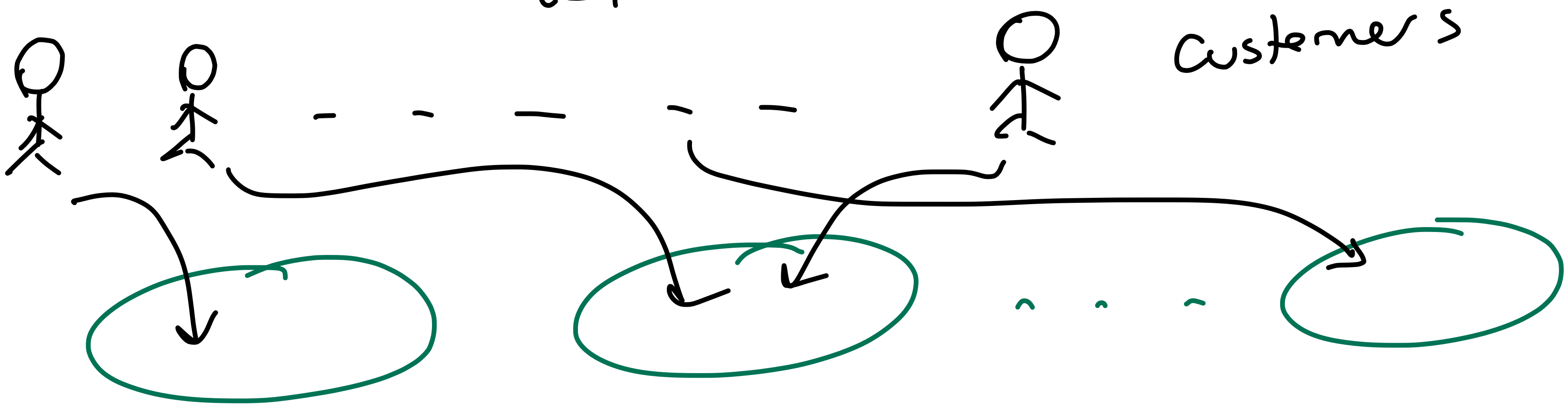
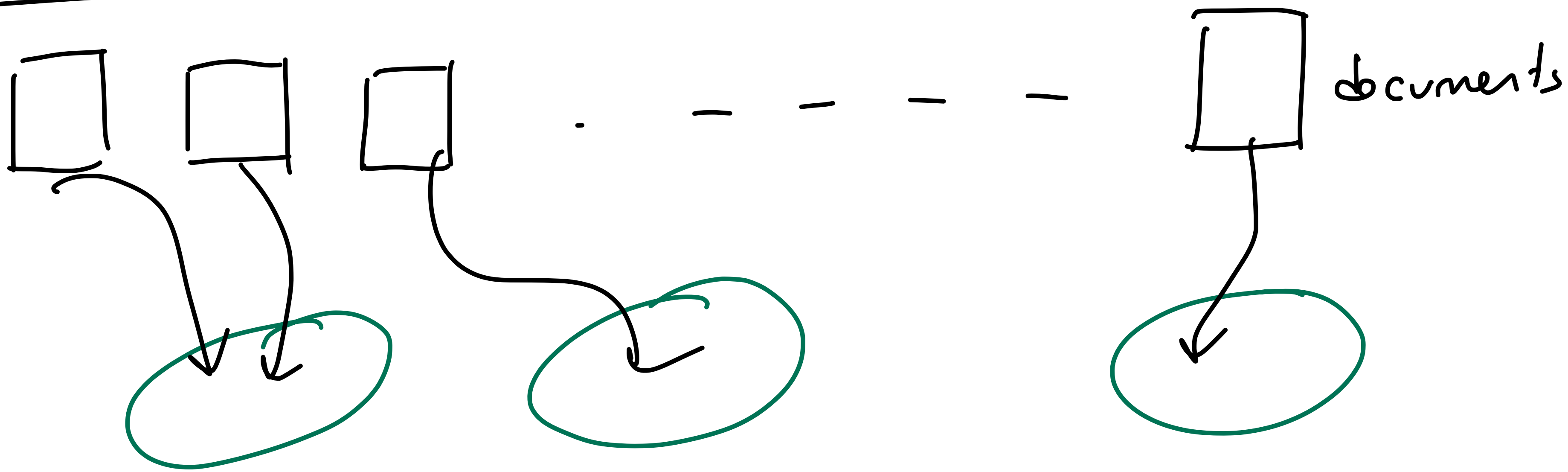
Unsupervised Learning

Clustering

$$\mathcal{X} = \{x_i\}_{i=1}^{N}$$

NO CLASS LABEL!





Supervised Learning

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

ith data point

ith label, ith output

classification

regression

Task: predicting whether a car is a family car or not

$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}$ $\begin{array}{l} \rightarrow \text{price} \\ \rightarrow \text{engine power} \end{array}$

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

car#1

car#2

$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{bmatrix}_{N \times D}$

$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}$

data matrix

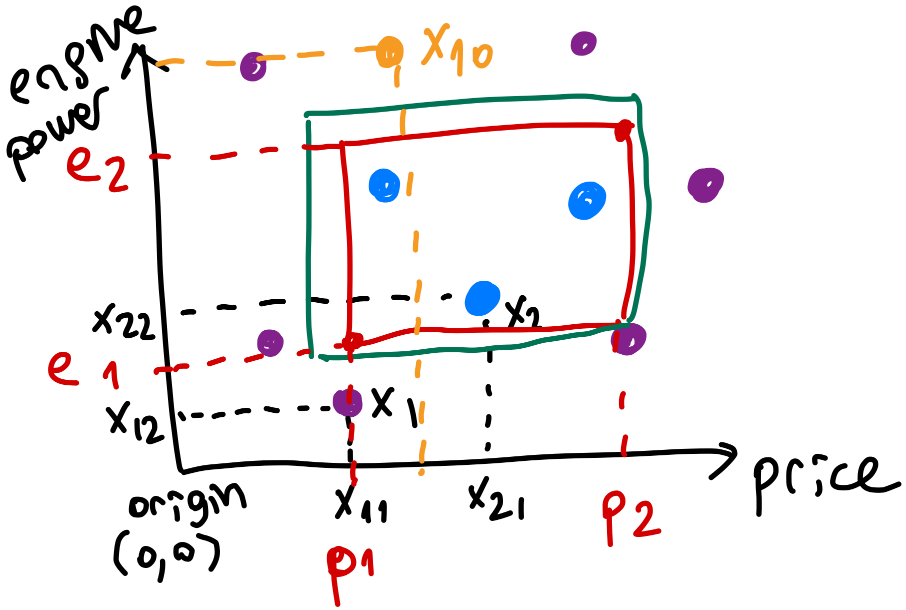
D → # of features (2)

N → # of data points

$\begin{bmatrix} x_{N+1,1} & x_{N+1,2} \end{bmatrix}$

$[\ ?\ ]$

label vector



● : family cars (positive)

● : other types of cars (negative)

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

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

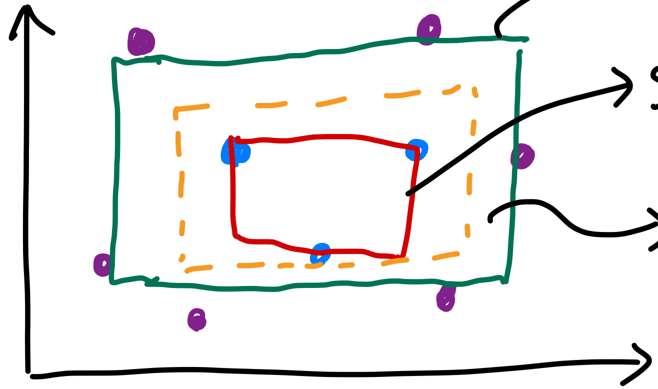
$$\underbrace{\text{RECTANGLES}}_{\text{model family}} \qquad \underbrace{\theta}_{\text{model parameters}} = \{e_1, e_2, p_1, p_2\}$$

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

LEARNING ⇒ finding the best θ ⇒ $e_1^*, e_2^*, p_1^*, p_2^*$

if $p_1 \leq x_{N+1,1} \leq p_2$ & $e_1 \leq x_{N+1,2} \leq e_2 \Rightarrow \hat{y}_{N+1} = 1$

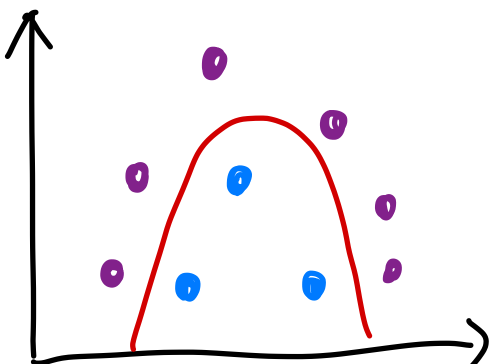
otherwise $\Rightarrow \hat{y}_{N+1} = 0$



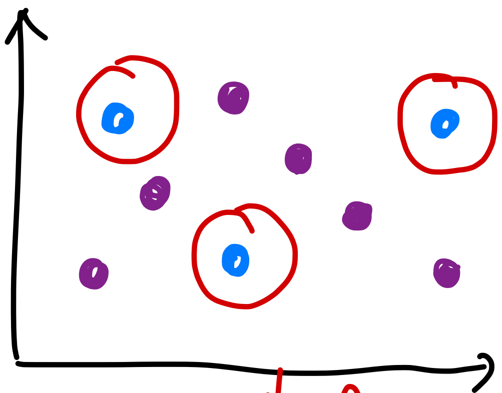
→ general solution

→ specific solution

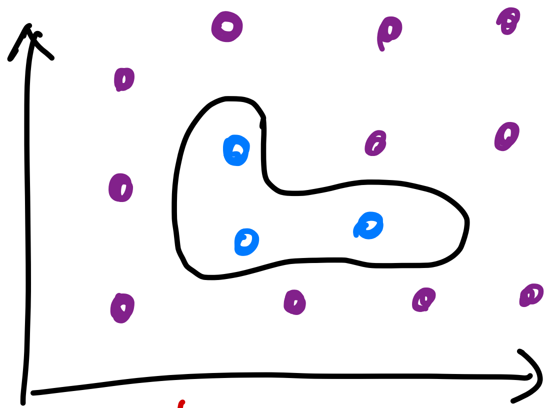
→ all rectangles in between are called "version space".



↓ set of second order polynomiels



↓ set of mixture of circles



↓ set of polygons

model complexity ↑

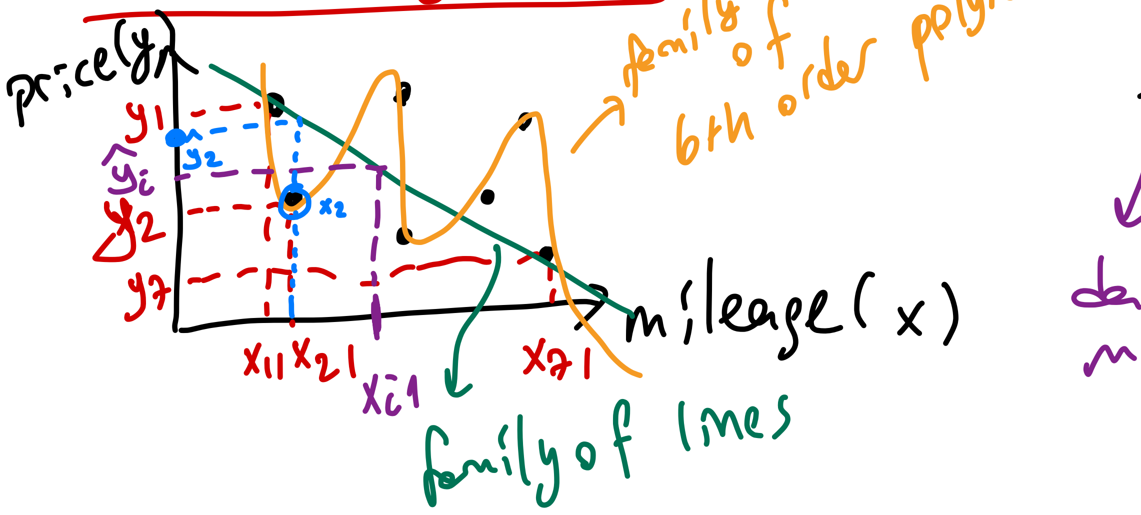
prediction performence

training ↑

test ↓

overfitting ⇒ at first both peformences will improve, after some point test performance will get worse.

Linear Regression



$$X = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{71} \end{bmatrix}_{7\times 1}$$

data matrix

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_7 \end{bmatrix}_{7\times 1}$$

output vector

$$\underbrace{\text{SET OF LINES}}_{\text{model family}}$$

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

$$\hat{y}_i = w_1 \cdot x_{i1} + w_0$$

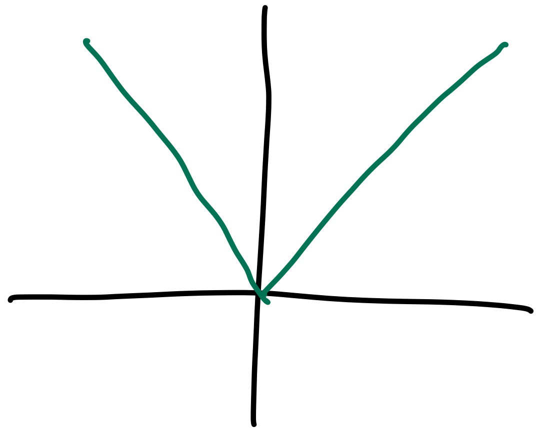
errors / observed values / predicted values

$$e_1 = y_1 - \hat{y}_1$$
$$e_2 = y_2 - \hat{y}_2$$
$$\vdots$$
$$e_N = y_N - \hat{y}_N$$

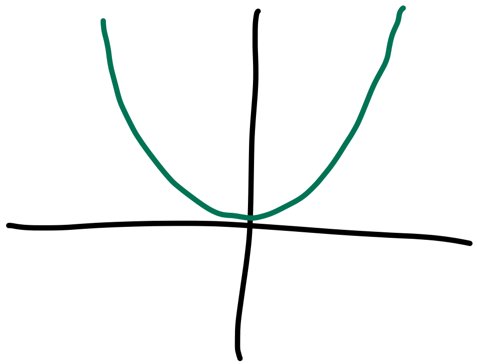
$$\hat{y}_1 = w_1 \cdot x_{11} + w_0$$
$$\hat{y}_2 = w_1 \cdot x_{21} + w_0$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\hat{y}_N = w_1\, x_{N1} + w_0$$

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

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



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



$$e_i = y_i - \hat{y}_i = y_i - [w_1 \cdot x_{i1} + w_0]$$

$$\text{minimize} \sum_{i=1}^{N} \left[y_i - [w_1 x_{i1} + w_0] \right]^2 = f(w_0, w_1)$$

training data.

$$\text{Error}(w_0, w_1 \mid \mathcal{X}) = \sum_{i=1}^{N} \left(y_i - (w_1 x_{i1} + w_0) \right)^2$$

$$\frac{\partial \text{Error}}{\partial w_0} = \sum_{i=1}^{N} \frac{\partial \left[y_i - (w_1 x_{i1} + w_0) \right]^2}{\partial w_0}$$

$$= \sum_{i=1}^{N} 2 \cdot \left[y_i - (w_1 x_{i1} + w_0) \right] \cdot (-1) = 0$$

$$\frac{\partial \text{Error}}{\partial w_1} = \sum_{i=1}^{N} \frac{\partial \left[y_i - (w_1 x_{i1} + w_0) \right]^2}{\partial w_1}$$

$$= \sum_{i=1}^{N} 2 \left[y_i - (w_1 . x_{i1} + w_0) \right] . (-x_{i1}) = 0$$

Exercise: Solve for w_0 & w_1.

$$w_0 = \left(\sum_{i=1}^{N} y_i / N \right) - w_1 \left(\sum_{i=1}^{N} x_{i1} / N \right)$$

$$w_1 = \frac{\sum_{i=1}^{N} x_{i1} y_i - \left(\sum_{i=1}^{N} x_{i1} / N \right) \left(\sum_{i=1}^{N} y_i / N \right) \cdot N}{\sum_{i=1}^{N} x_{i1}^2 - N \left(\sum_{i=1}^{N} x_{i1} / N \right)^2}$$