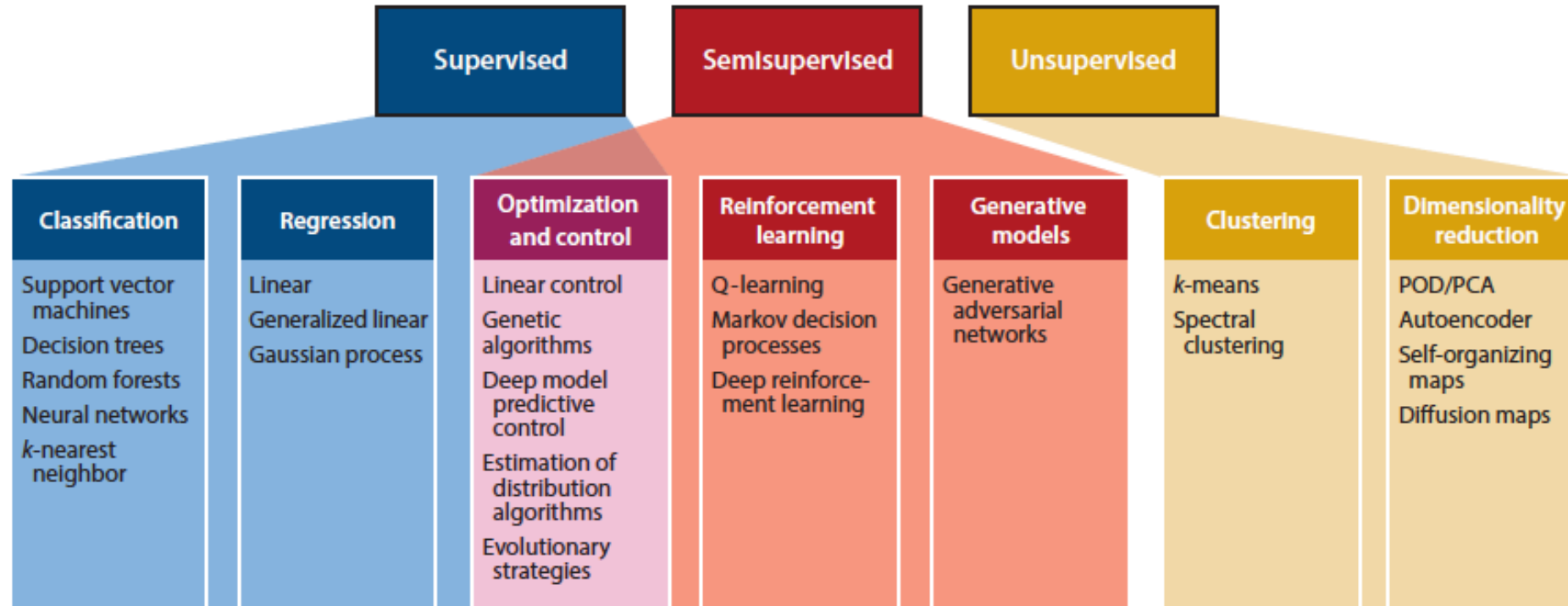


# FUNDAMENTOS DE MACHINE LEARNING: CLASSIFICAÇÃO E REGRESSÃO

Fevereiro de 2021

Ref.: Cap. 4 e 5 do livro texto (Regression and Model Selection /  
Clustering and Classification)

# Machine Learning: uma visão panorâmica de métodos, algoritmos e modelos

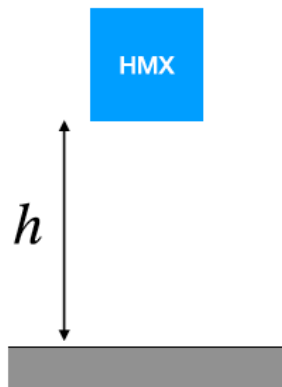


**Figure 1**

Machine learning algorithms may be categorized into supervised, unsupervised, and semisupervised, depending on the extent and type of information available for the learning process. Abbreviations: PCA, principal component analysis; POD, proper orthogonal decomposition.

Machine Learning : aprender com os dados  
porque, como e para que?  
**EXTRAPOLAR...**

## Example: Sensitivity of energetic materials



Does it explode or not when you drop it from height  $h$ ?

Height (cm)	Results
40.5	E E E E E E E E E E
36.0	E N E E E E N E E E
32.0	E E N E E E N E N E
28.5	N E N N E N N N N E N
25.5	N N N N N N N E N N N
22.5	N N N N N N N N N N N

*Data from L. Smith, "Los Alamos National Laboratory explosives orientation course: Sensitivity and sensitivity tests to impact, friction, spark and shock," Los Alamos National Lab, NM (USA), Tech. Rep., 1987*

Construir um modelo que permita “melhor compreender o que os  
: dados informam” .

- Simbolicamente o modelo

$$\underbrace{\mathcal{Y}}_{saídas} = \mathcal{M}(\underbrace{\mathcal{X}}_{entradas})$$

onde (neste contexto) a entrada é a altura e a saída a **possibilidade** de haver explosão.

Problema direto (PD) : qual o valor da saída para uma entrada específica? (de fato, qual a possibilidade de explodir o artefato (análise de risco) segundo o modelo a ser construído).

Problema inverso (PI) : qual a menor altura segura?

As incertezas nos dados e na construção do modelo, ambos os problemas poderiam ser formulados sob uma ótica probabilística...

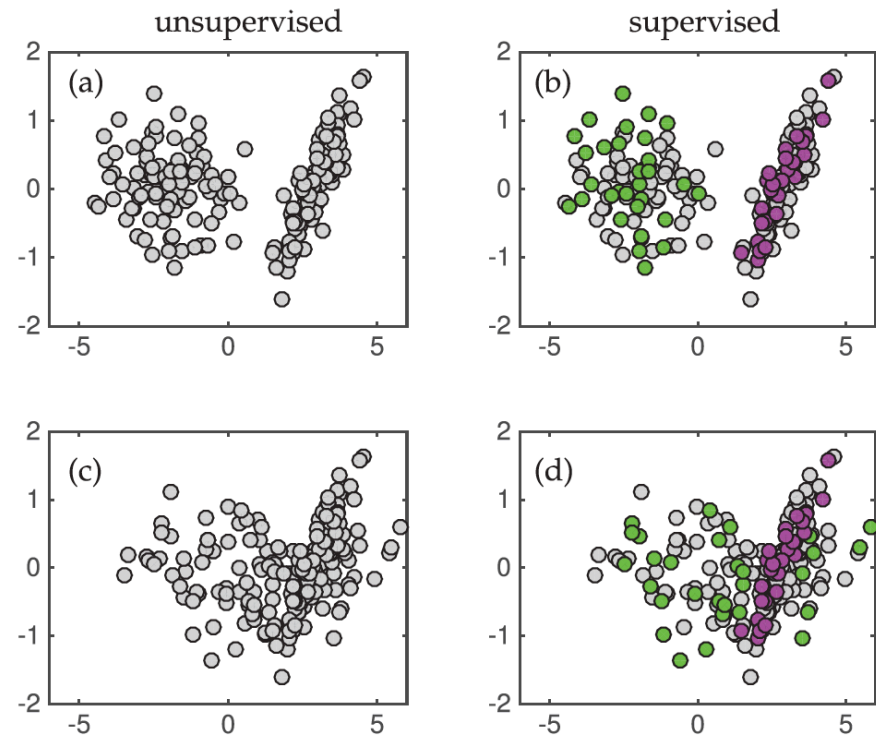
- PD (dado a natureza do problema (note que a saída não é uma variável contínua.. um classificador ... estimar a chance de pertencer a uma classe)

$$P(\mathcal{Y}|\mathcal{X})$$

- PI ... encontrar o valor mínimo (ótimo) sob incertezas... tolerância ao risco

$$P(\mathcal{Y} == E) < \delta$$

# Aprender através dos dados : os dois paradigmas



**Figure 5.7** Illustration of unsupervised versus supervised learning. In the left panels (a) and (c), unsupervised learning attempts to find clusters for the data in order to classify them into two groups. For well separated data (a), the task is straightforward and labels can easily be produced. For overlapping data (c), it is a very difficult task for an unsupervised algorithm to accomplish. In the right panels (b) and (d), supervised learning provides a number of labels: green balls and magenta balls. The remaining unlabeled data is then classified as green or magenta. For well separated data (b), labeling data is easy, while overlapping data presents significant challenge.

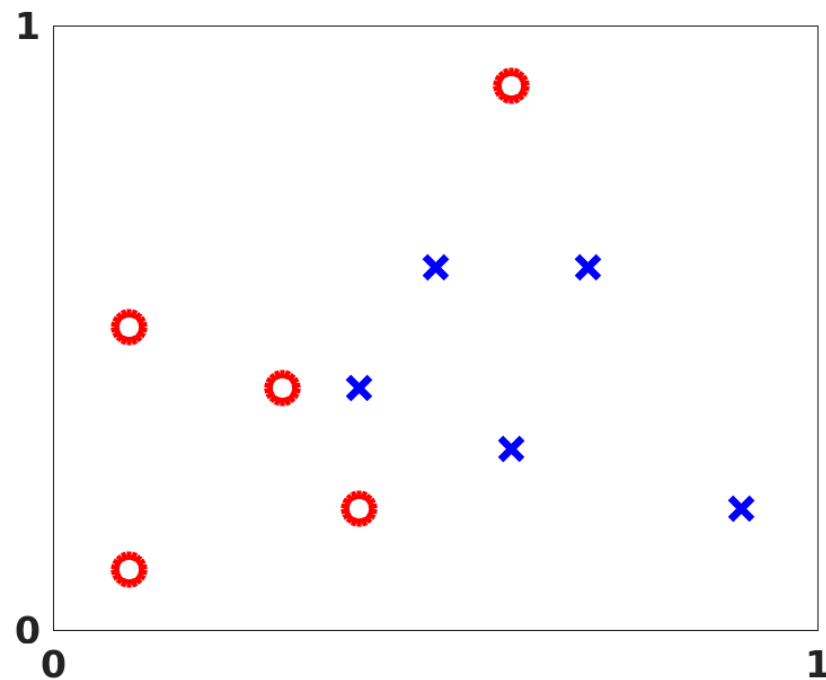
# Aprendizado Supervisionado

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$

$$y_{1:n} = (y_1, \dots, y_n)$$

Dados rotulados ou anotados (labeled)

# Um modelo para classificação

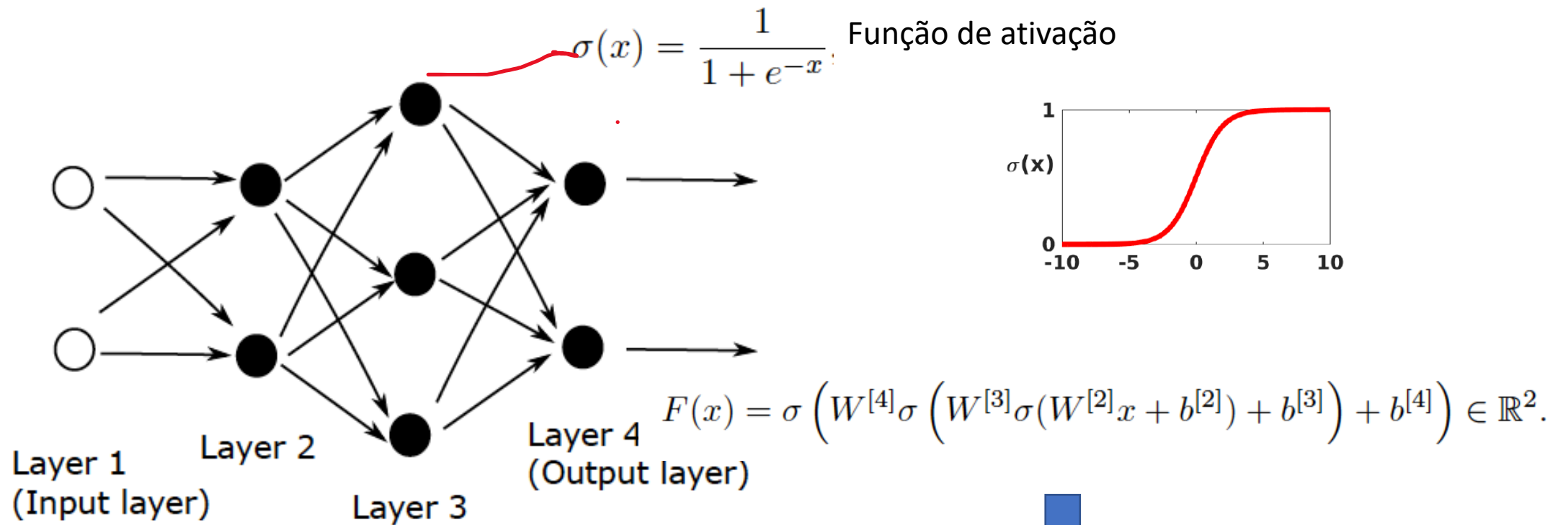


Catherine F. Highamy, Desmond J. Highamz,  
Deep Learning: An Introduction for Applied Mathematicians, SIAM  
REVIEW 2019 SIAM. Published by SIAM under the terms  
Vol. 61, No. 4, pp. 860–891

“To be concrete, consider the set of points shown in Figure . This shows labeled data some points are in category A, indicated by circles, and the rest are in category B, indicated by crosses. For example, the data may show oil drilling sites on a map, where category A denotes a successful outcome. Can we use this data to categorize a newly proposed drilling site? Our job is to construct a mapping that takes any point in  $\mathbb{R}^2$  and returns either a circle or a cross. Of course, there are many reasonable ways to construct such a mapping.”



# Diversidade de possibilidades...



modelo determinado pelas matrizes  $W^i$  e vetores  $b^i$ . (certamente também pela arquitetura da rede)

Racionalizando e treinando a rede (não se trata de um classificador clássico neste exemplo)

- O treinamento consiste em encontrar os 23 parâmetros que definem o modelo. Isto pode ser feito, no contexto de aprendizado supervisionado, através do estabelecimento de um problema de otimização (minimização da distância entre dados e valores obtidos com o modelo , a rede neural).

$$\text{Cost} \left( W^{[2]}, W^{[3]}, W^{[4]}, b^{[2]}, b^{[3]}, b^{[4]} \right) = \frac{1}{10} \sum_{i=1}^{10} \frac{1}{2} \| y(x^{\{i\}}) - F(x^{\{i\}}) \|_2^2.$$



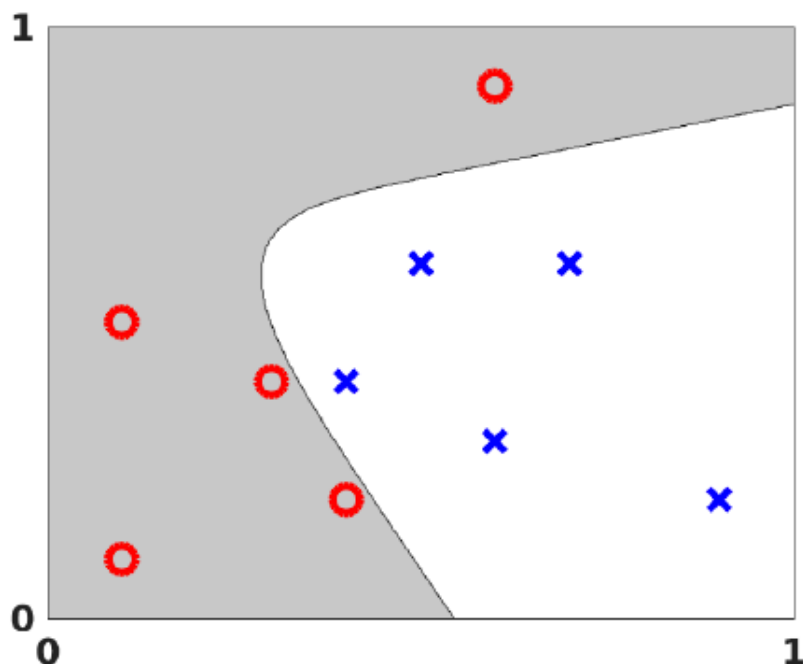
dados

## Racional do classificador (um pouco de abstração)

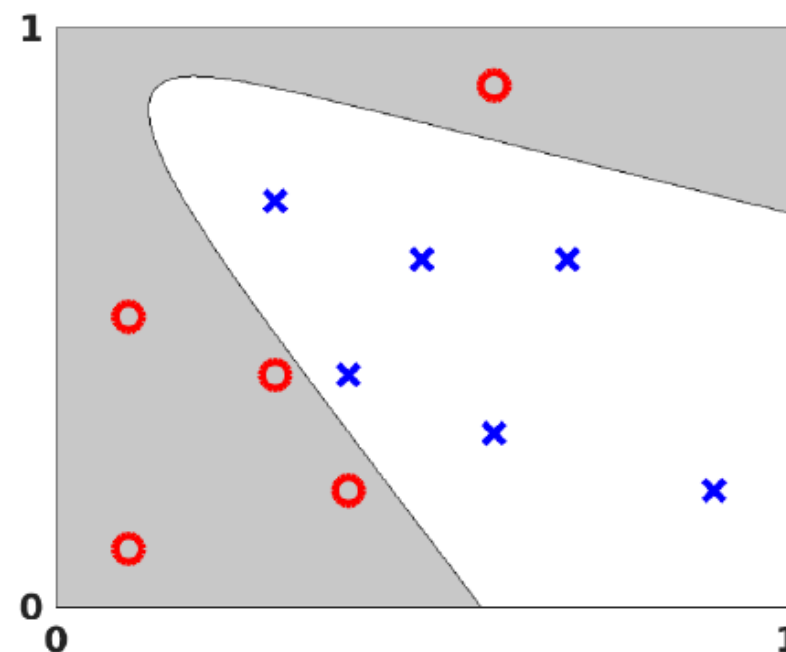
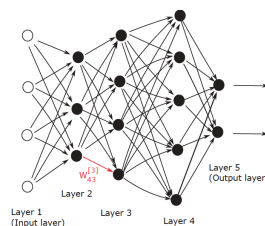
rameters. We will require  $F(x)$  to be close to  $[1, 0]^T$  for data points in category A and close to  $[0, 1]^T$  for data points in category B. Then, given a new point  $x \in \mathbb{R}^2$ , it would be reasonable to classify it according to the largest component of  $F(x)$ ; that is, category A if  $F_1(x) > F_2(x)$  and category B if  $F_1(x) < F_2(x)$ , with some rule to break ties. This requirement on  $F$  may be specified through a *cost function*. Denoting the

$$y(x^{\{i\}}) = \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } x^{\{i\}} \text{ is in category A,} \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } x^{\{i\}} \text{ is in category B.} \end{cases}$$

Usando um algoritmo de otimização de sua plataforma (neste caso a função lsqnonlin do Matlab) e extrapolando com o modelo treinado



Com mais dados



e se uma outra arquitetura fosse utilizada?  
Balanço entre maior capacidade de  
representação dos dados e dificuldade em  
treinar

