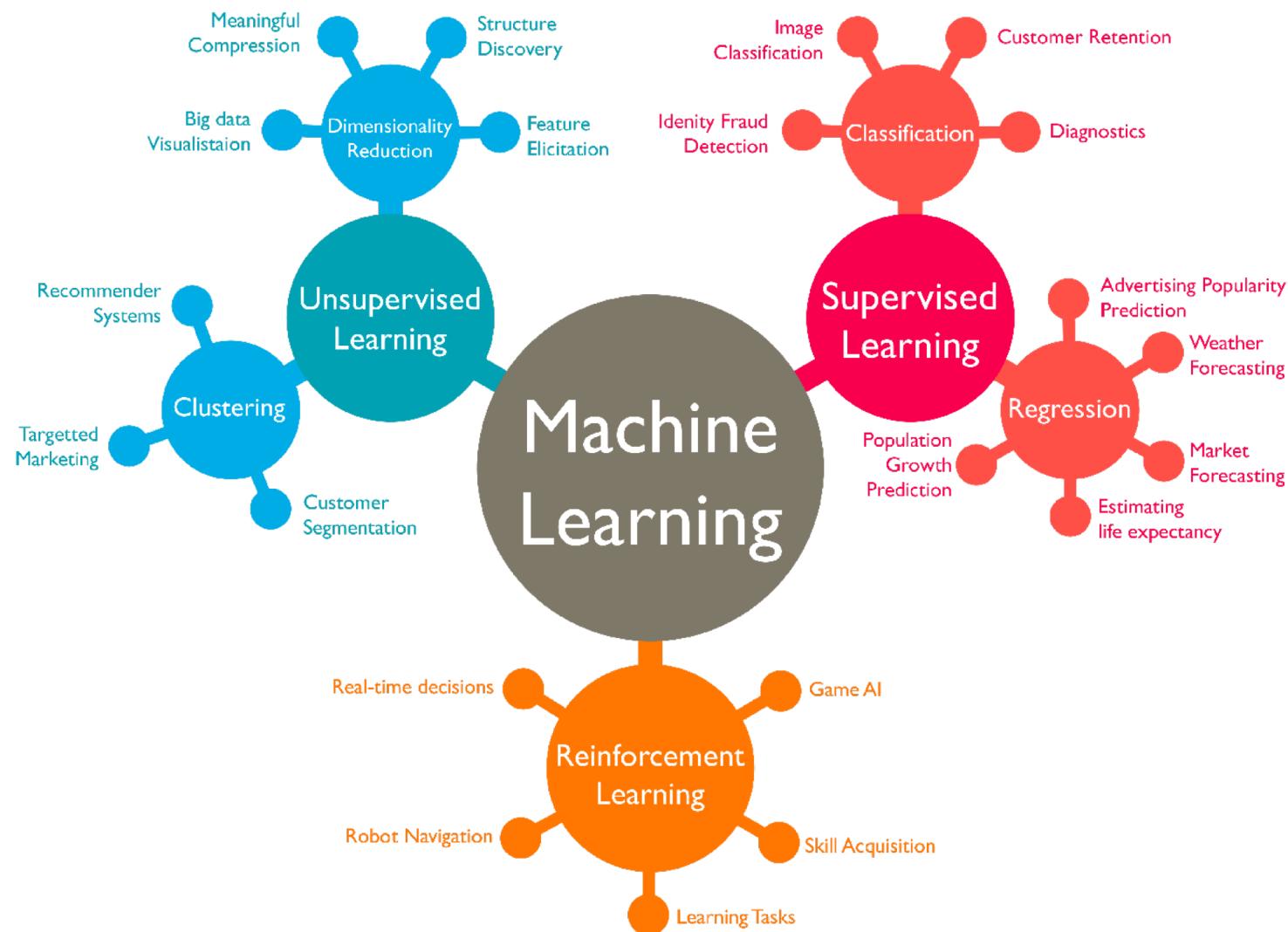




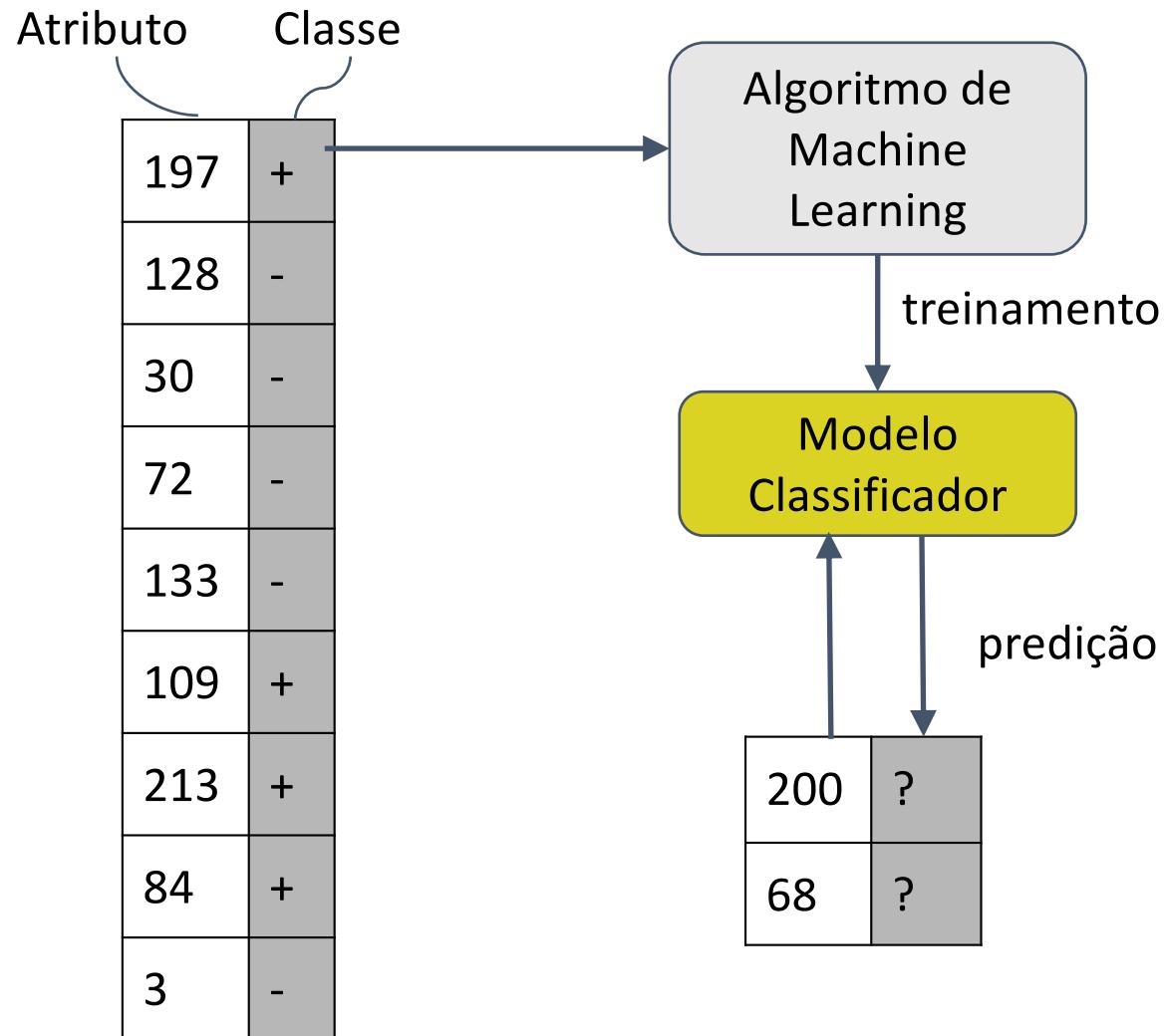
INSIGHT

Data Science Laboratory
Federal University of Ceará

www.insightlab.ufc.br



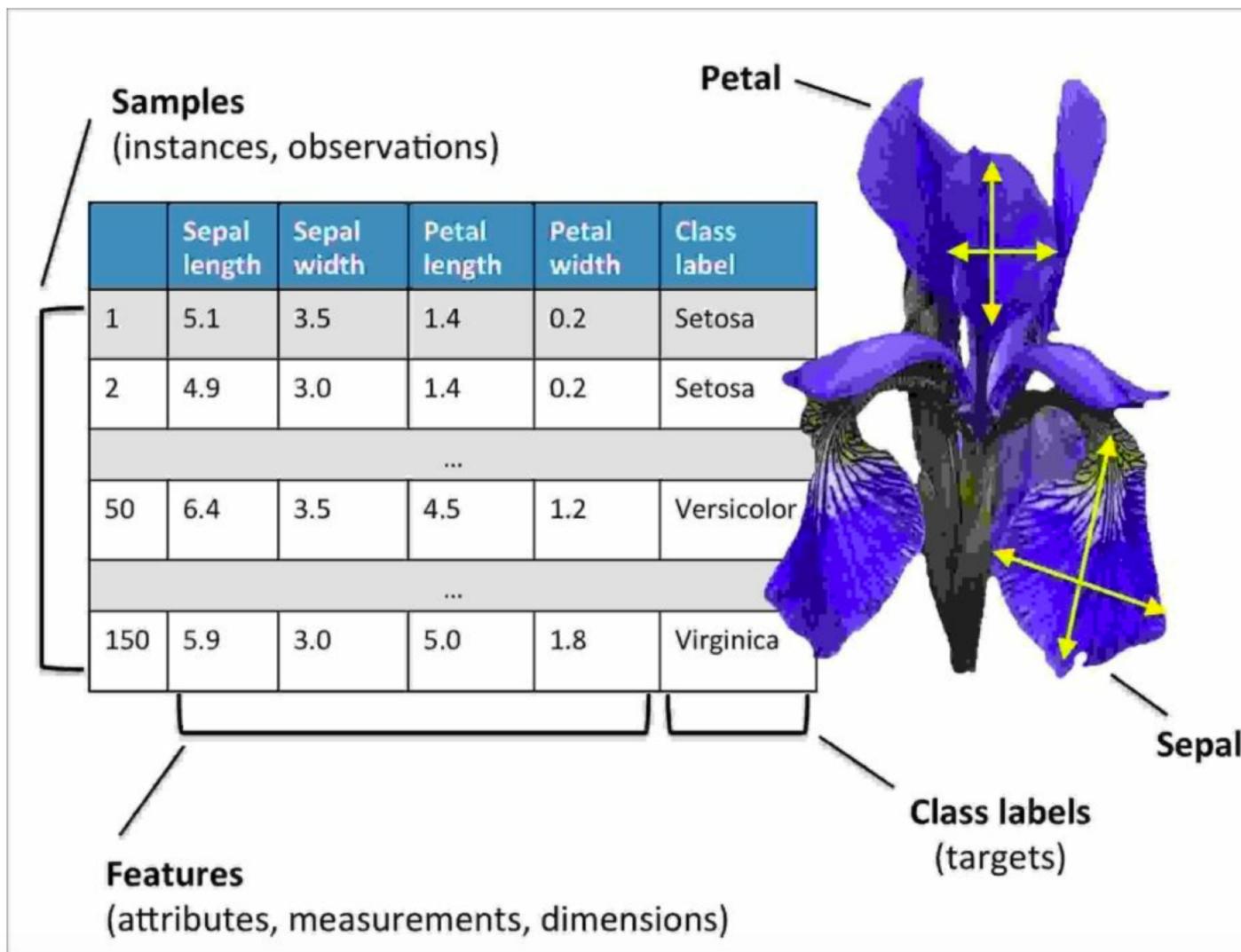
Como os algoritmos "aprendem" ?

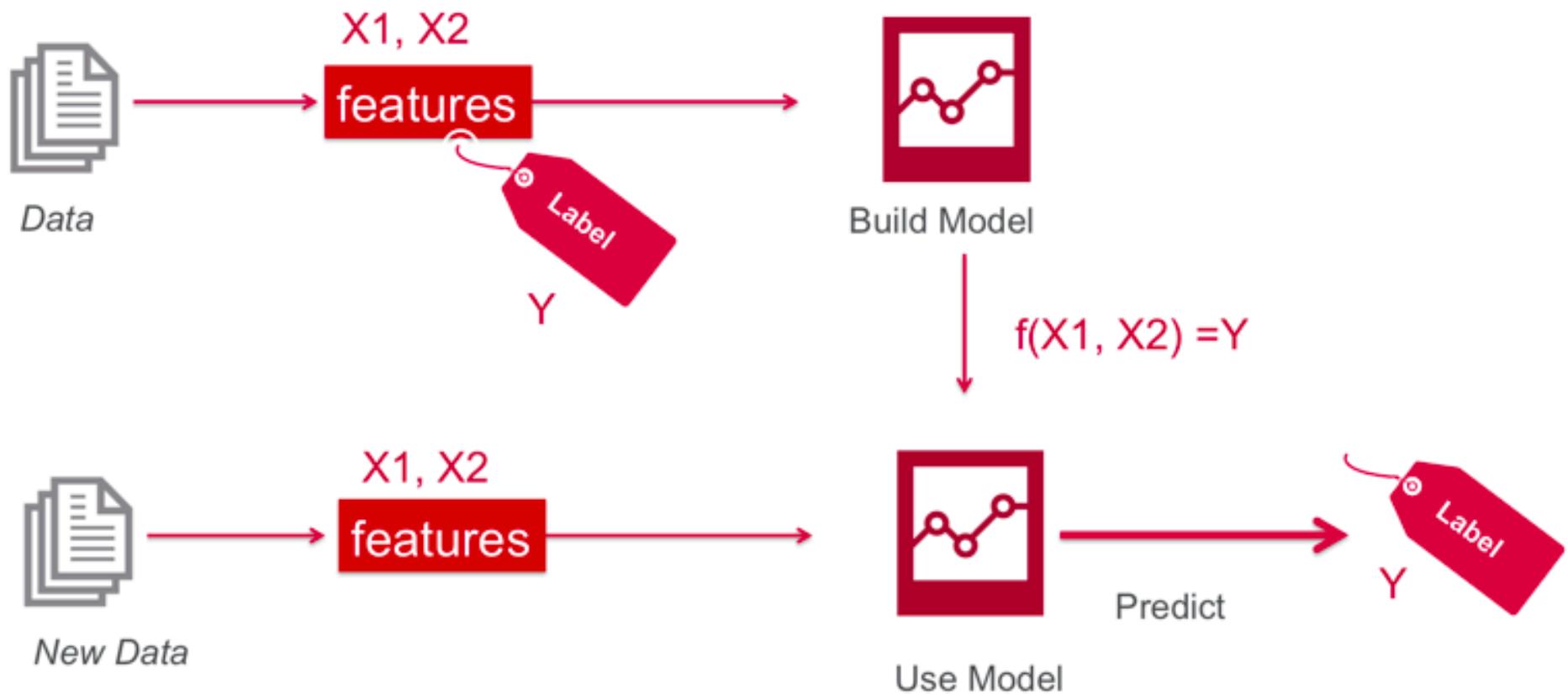


Representação do dado é fundamental

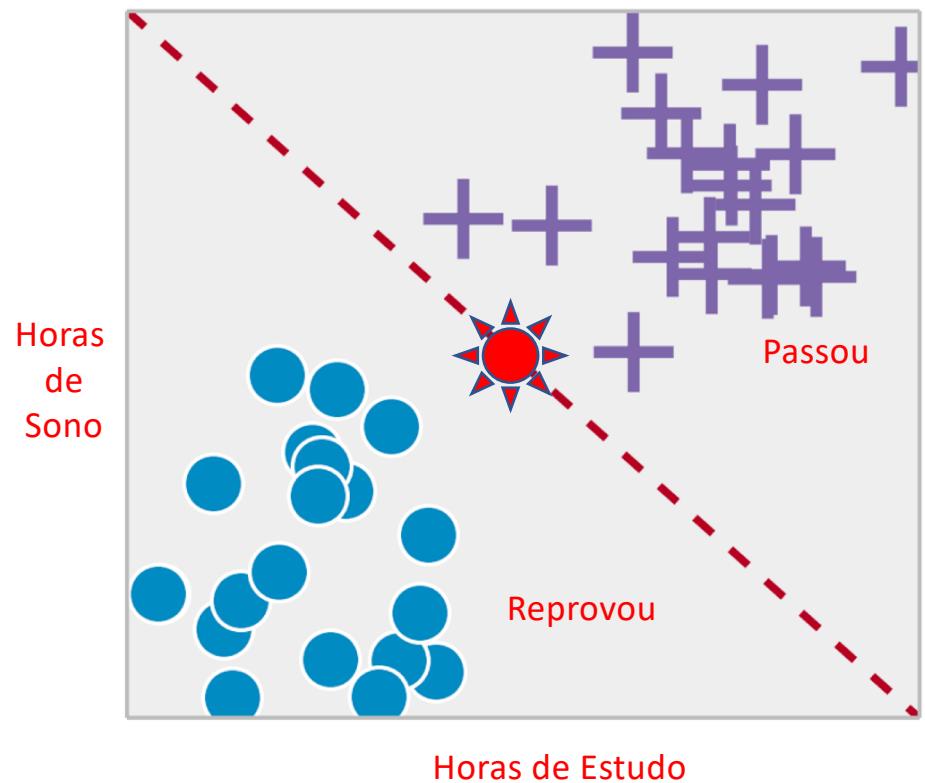
	Feature	Classe
197	+	11000101
128	-	10000000
30	-	00011110
72	-	01001000
133	-	10000101
109	+	01101101
213	+	11010101
84	+	01010100
3	-	00000011

200	?	11001000
68	?	01000100

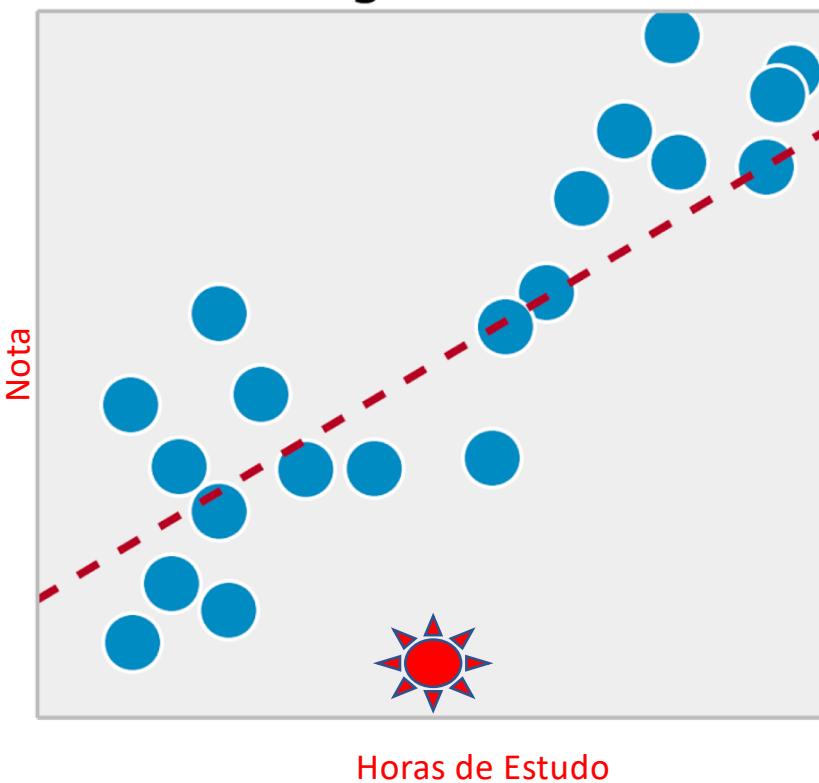




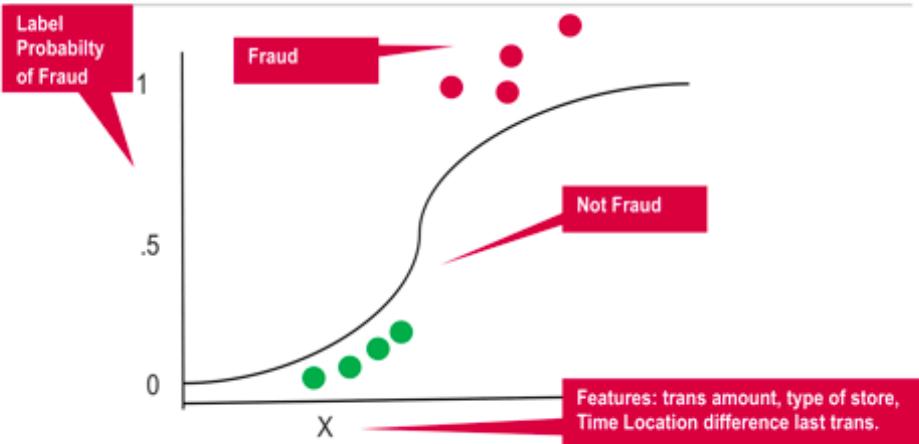
Classification



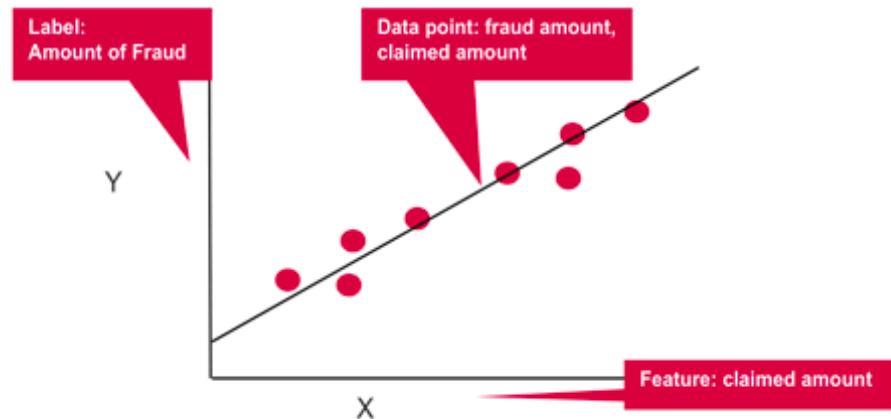
Regression



Credit Card Fraud Logistic Regression Classification Example

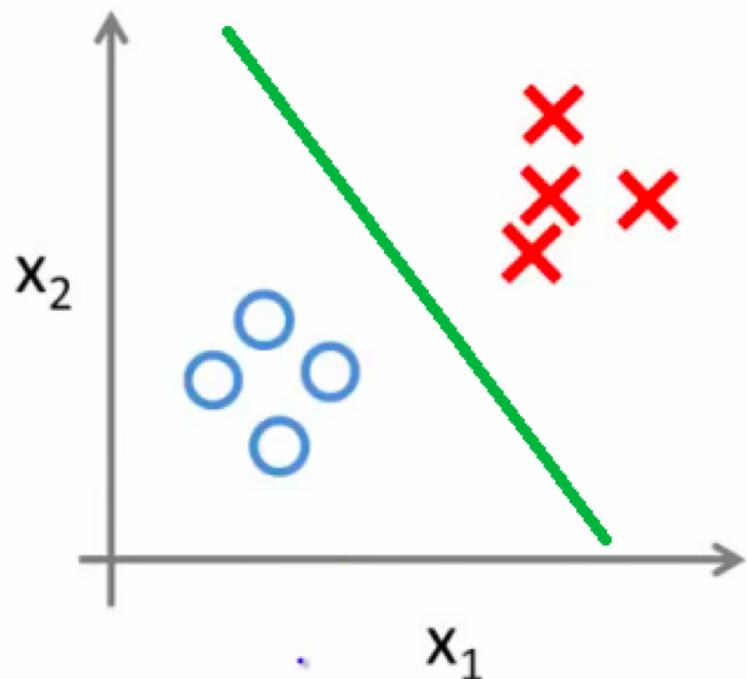


Car Insurance Fraud Regression Example

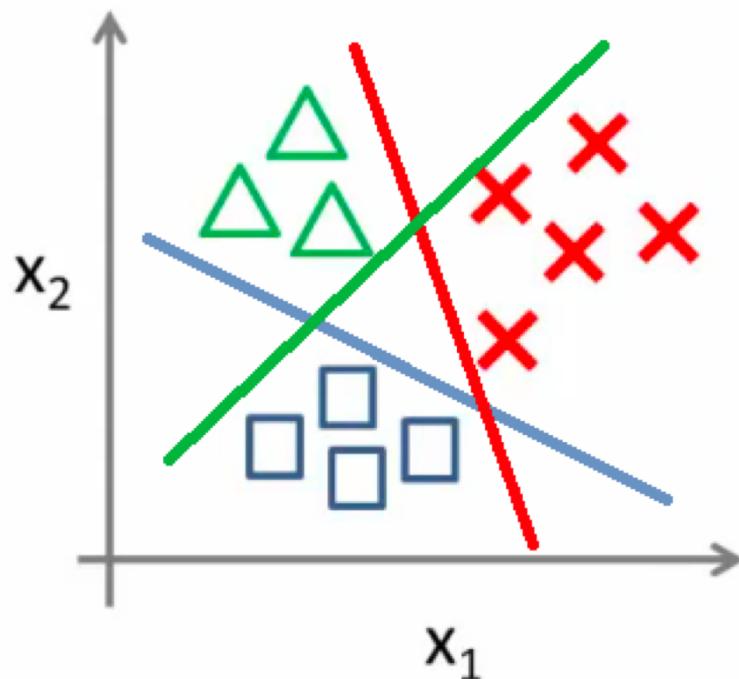


$$\text{AmntFraud} = \text{intercept} + \text{coeff} * \text{claimedAmnt}$$

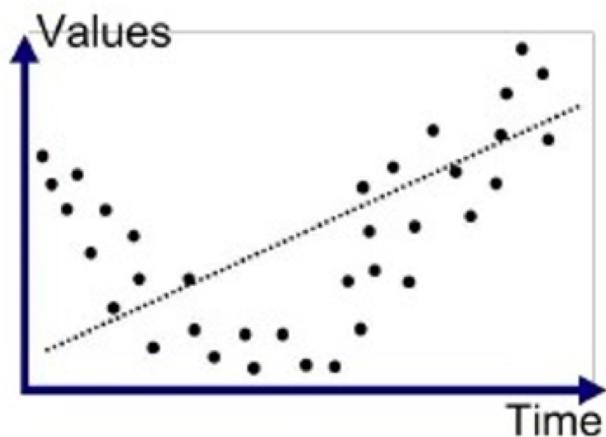
Binary classification:



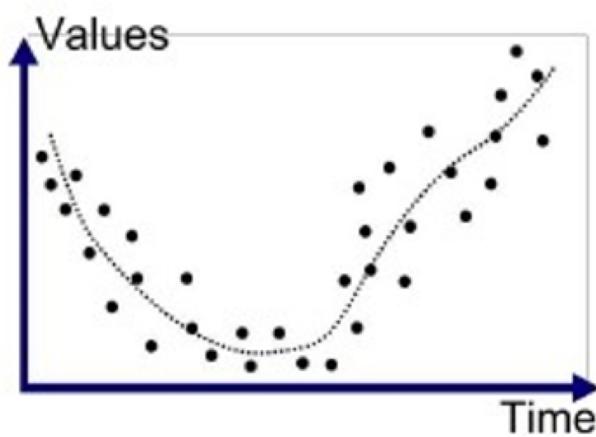
Multi-class classification:



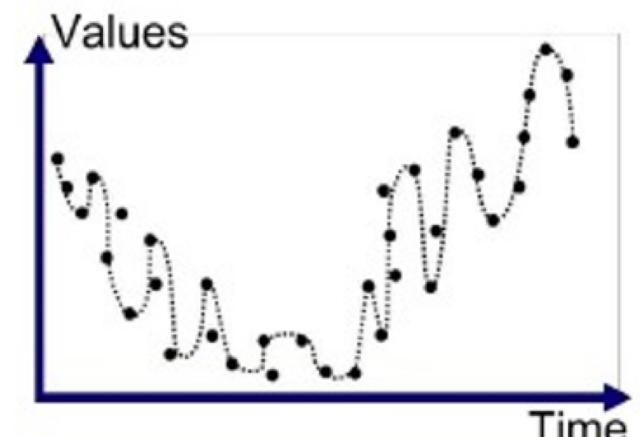
Ajustando um Modelo aos Dados



Underfitted



Good Fit/R robust



Overfitted

MACHINE LEARNING GENERALIZATION

FINDING THE PERFECT FIT

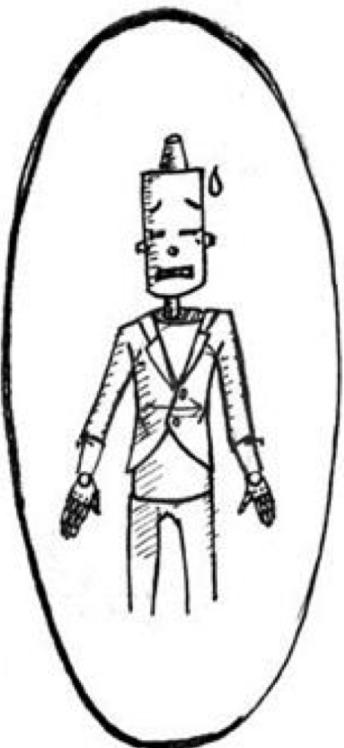
UNDERFIT

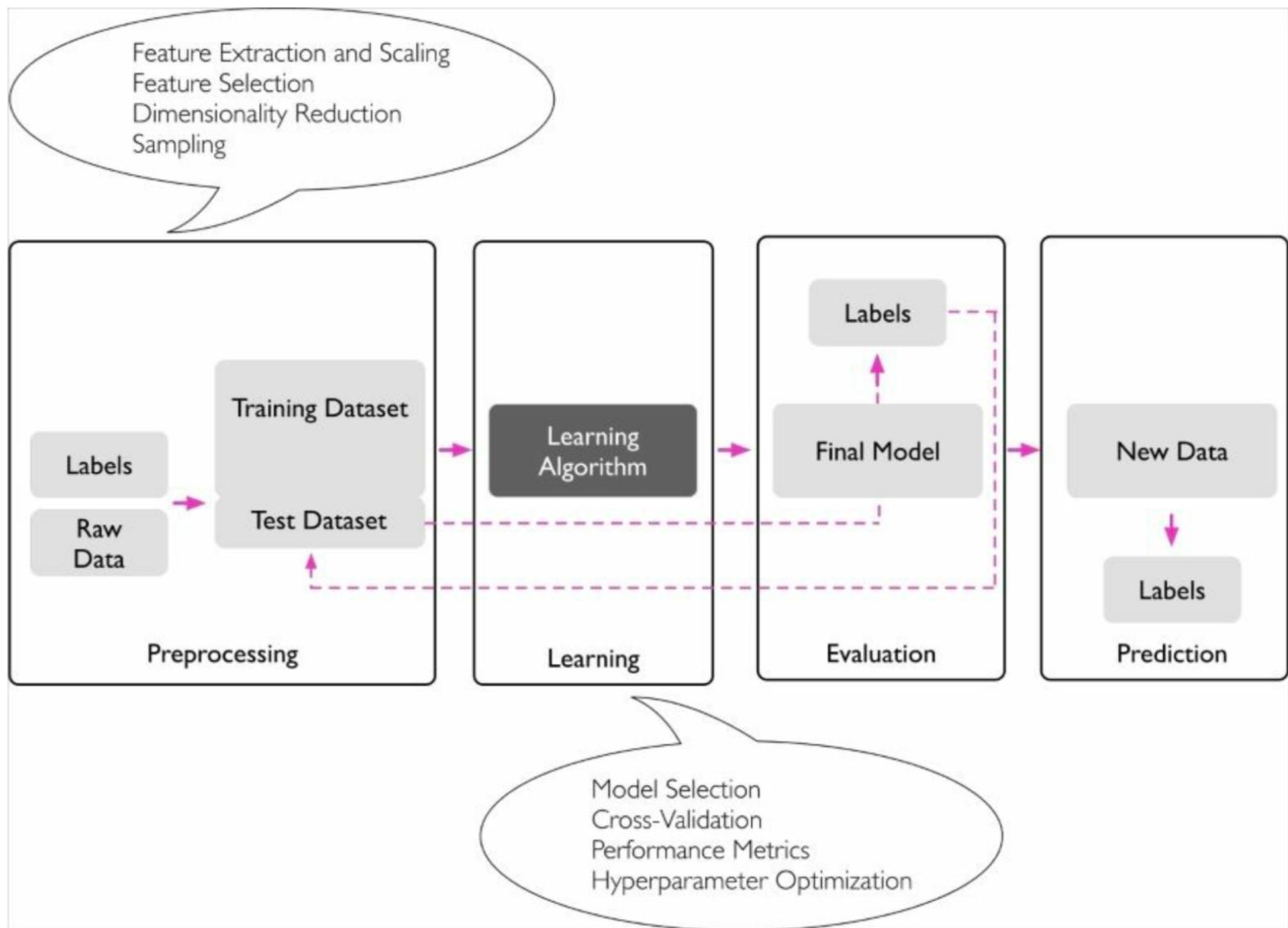


GOLDILOCKS ZONE

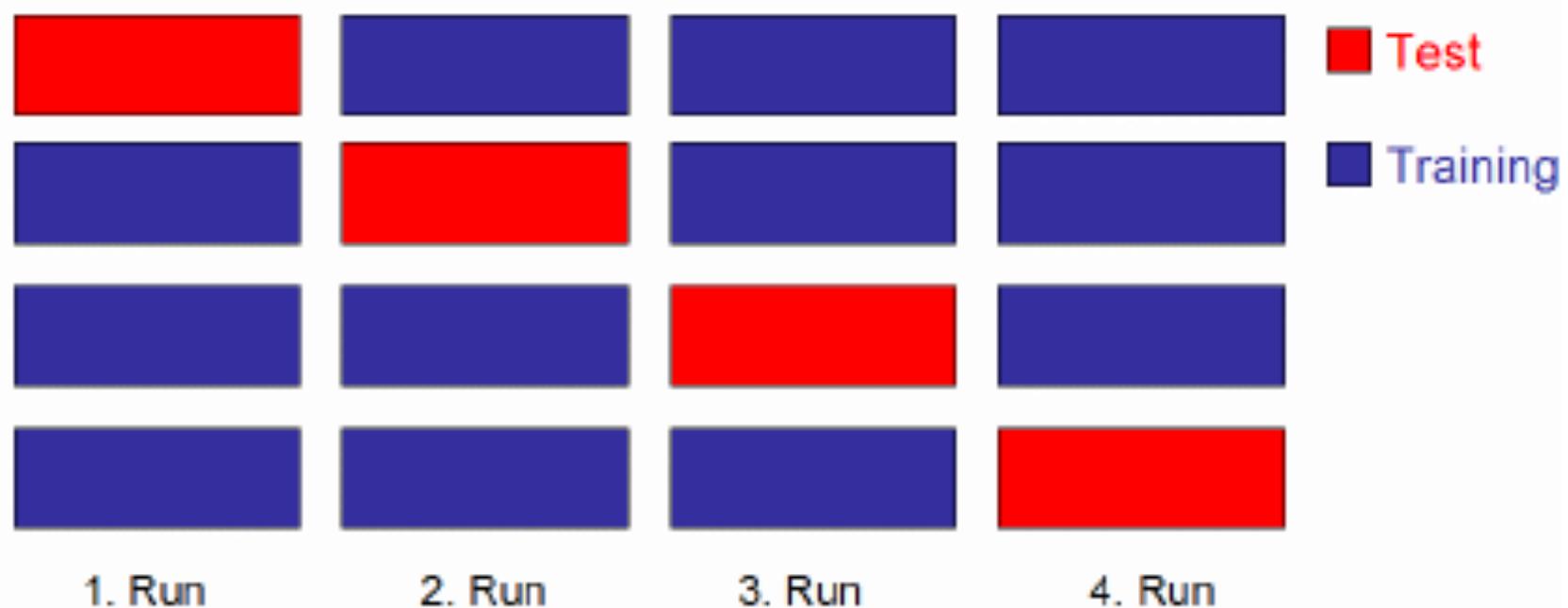


OVERFIT





k-fold Cross Validation Scheme (k=4)



Clusterização

