Oficina Gratuíta de Machine Learning

Café Coworking - Tecnopuc

http://goo.gl/BgoJmm

Facilitadores

Camila

Henrique

João

Jonatas

KUNUMI

Grupo de **Inteligência Artificial** na Saúde

RENNER

Edições Anteriores (259 participantes)

poa.hub

PUCRS

(PPGCC e PPGB)

TECNOPUC EXPERIENCE

memed

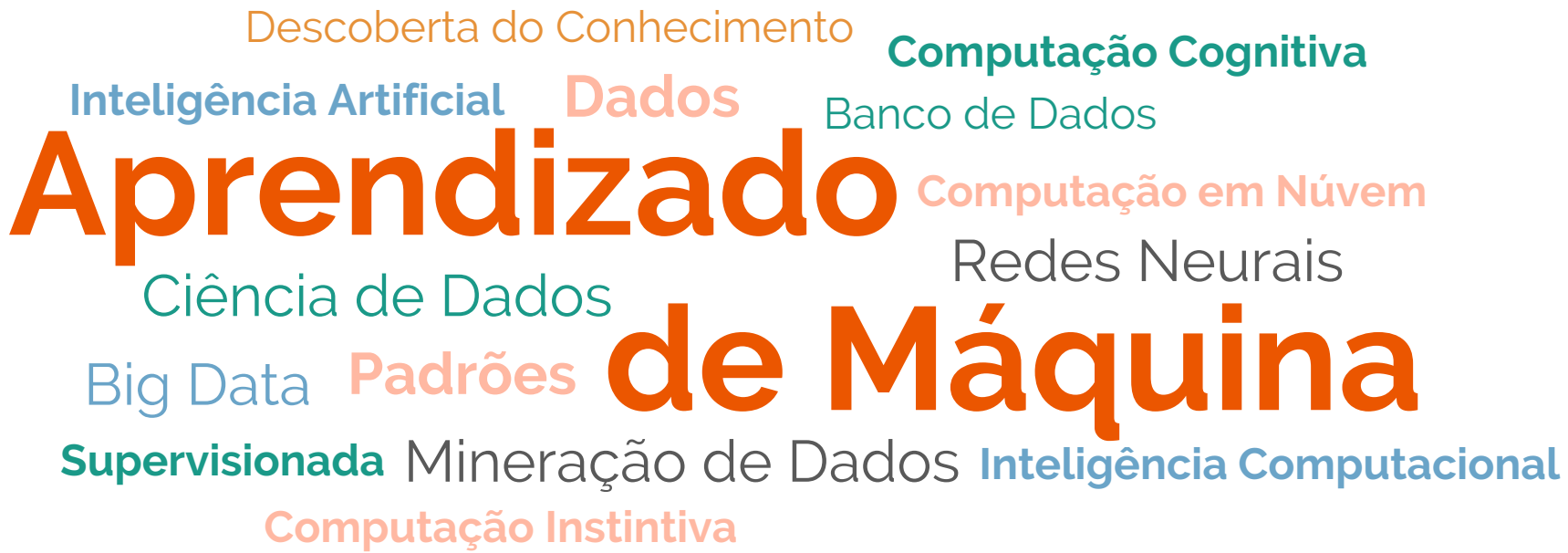
Google I/O 18 Extended

ft18

TECNOPUC

(Setembro e Dezembro)

Sopa de Letrinhas



Ferramentas

Machine Learning

scikit learn

K + TensorFlow

jupyter

Big Data

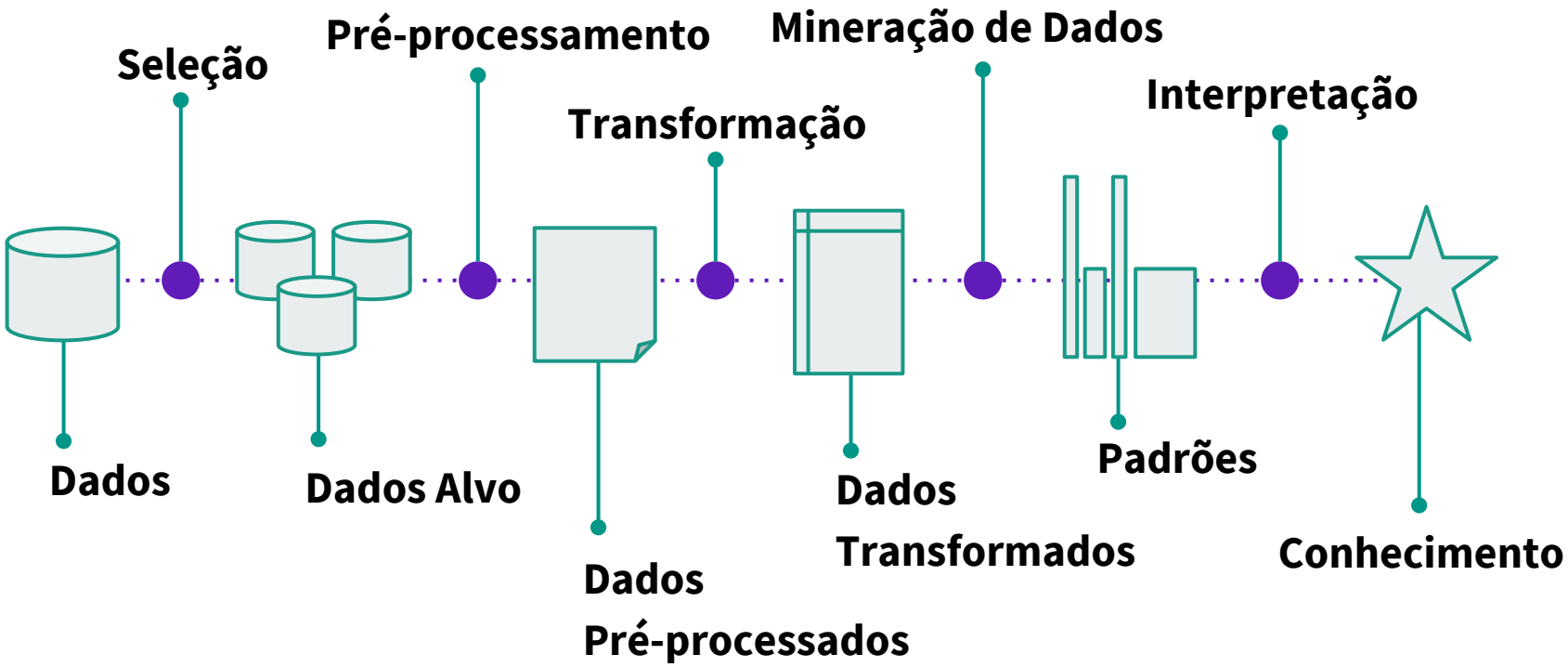
APACHE Spark™

hadoop

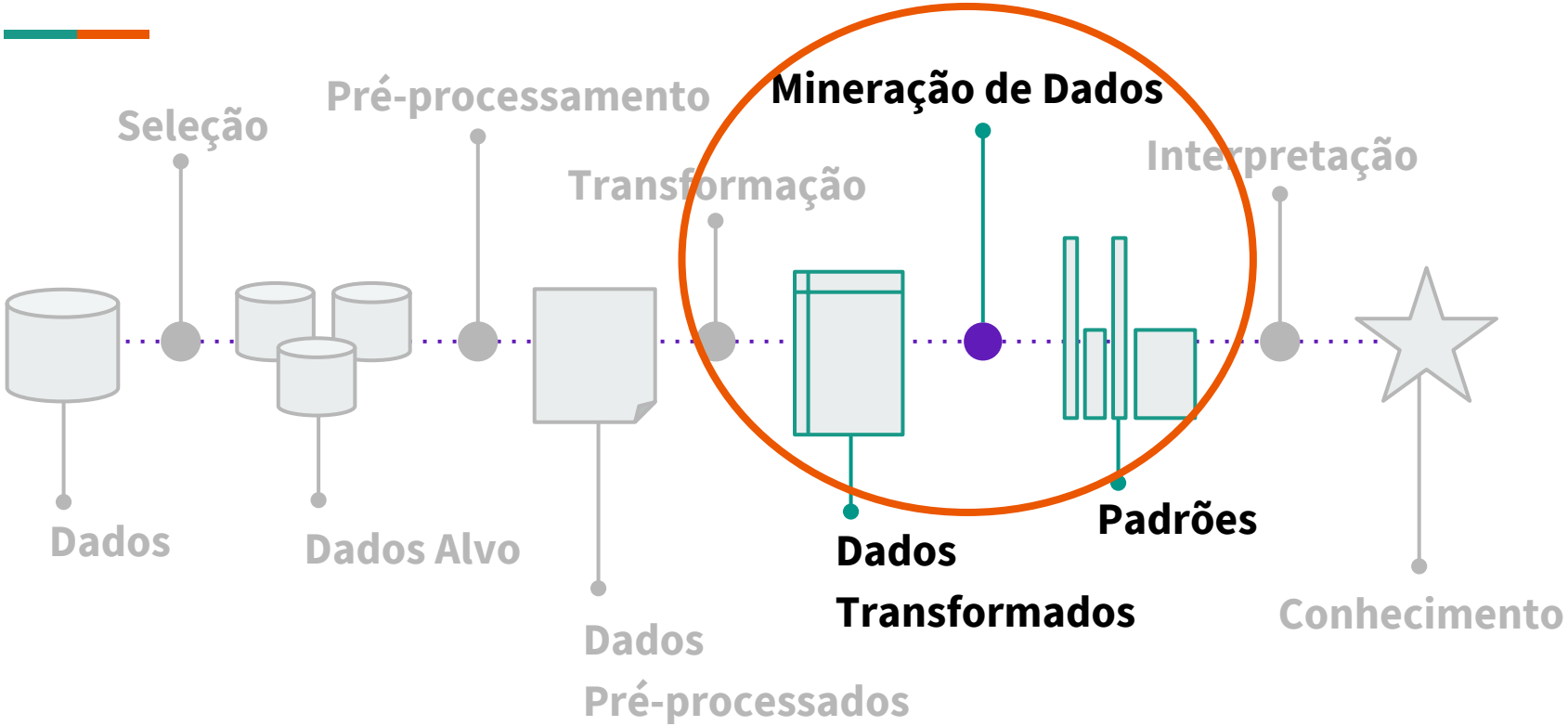
mongoDB

HIVE

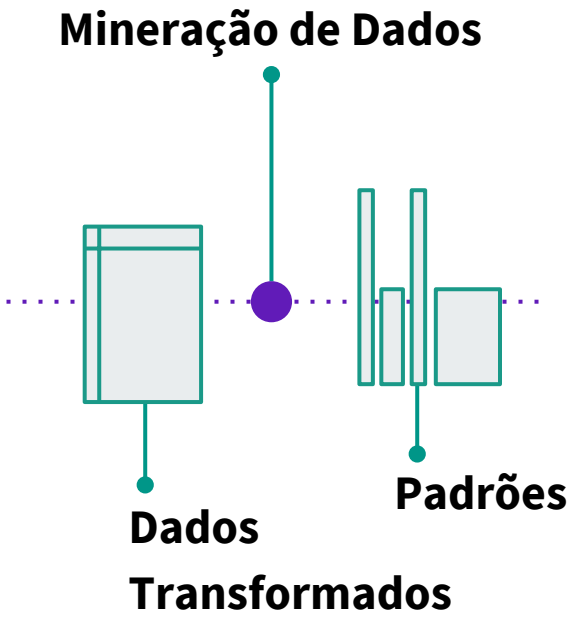
Descoberta de Conhecimento



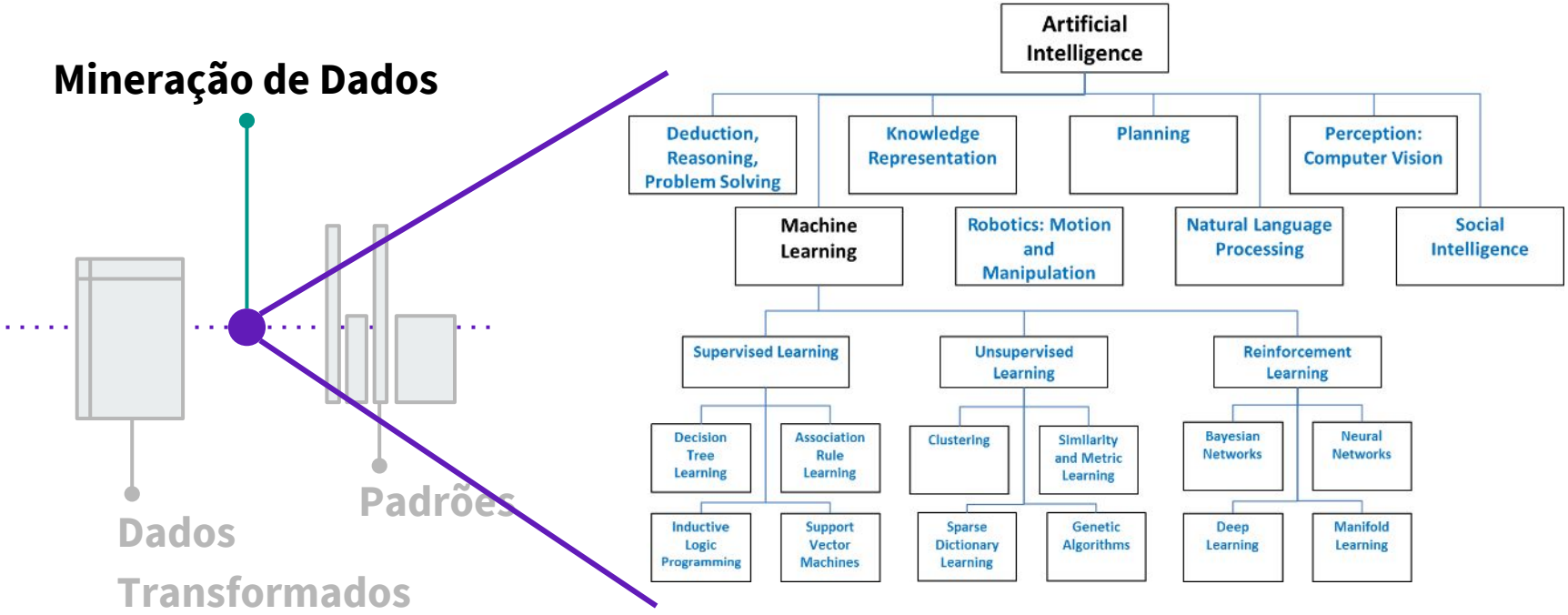
Descoberta de Conhecimento



Mineração de Dados



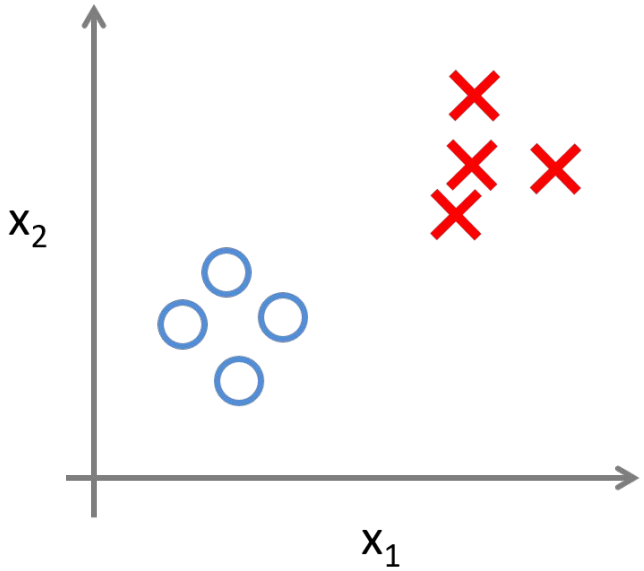
Aprendizado de Máquina



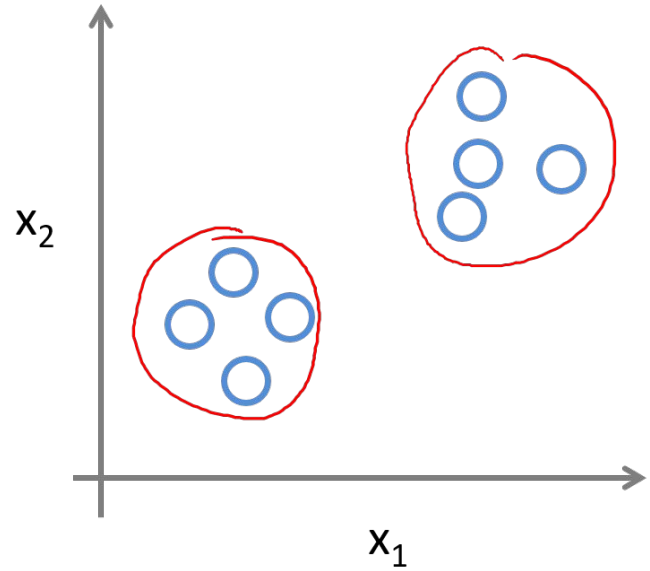
Tipos de Aprendizagem

Aprendizado de Máquina

Supervised Learning

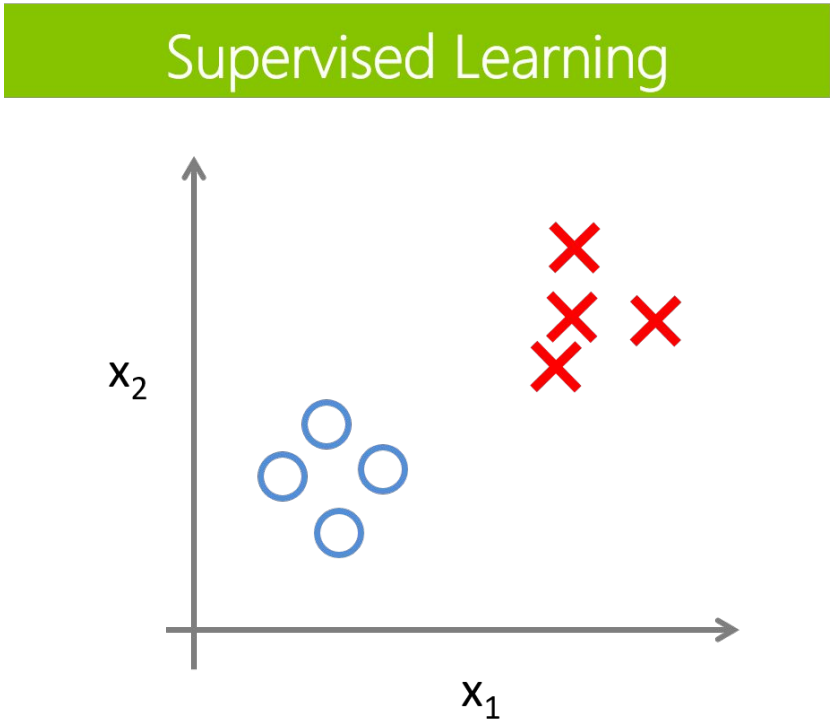


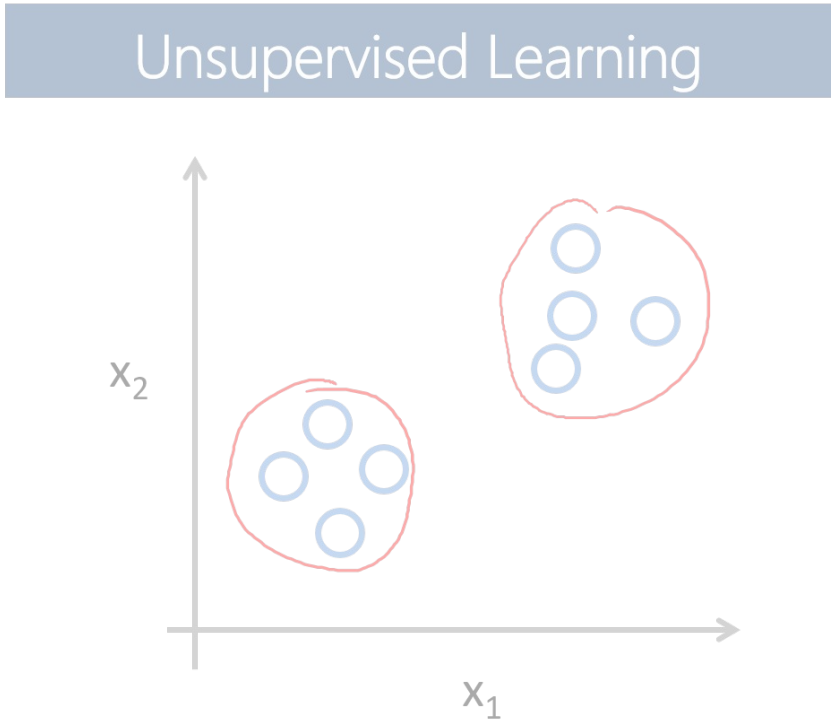
Unsupervised Learning



Imagens: https://leonardoaraujosantos.gitbooks.io

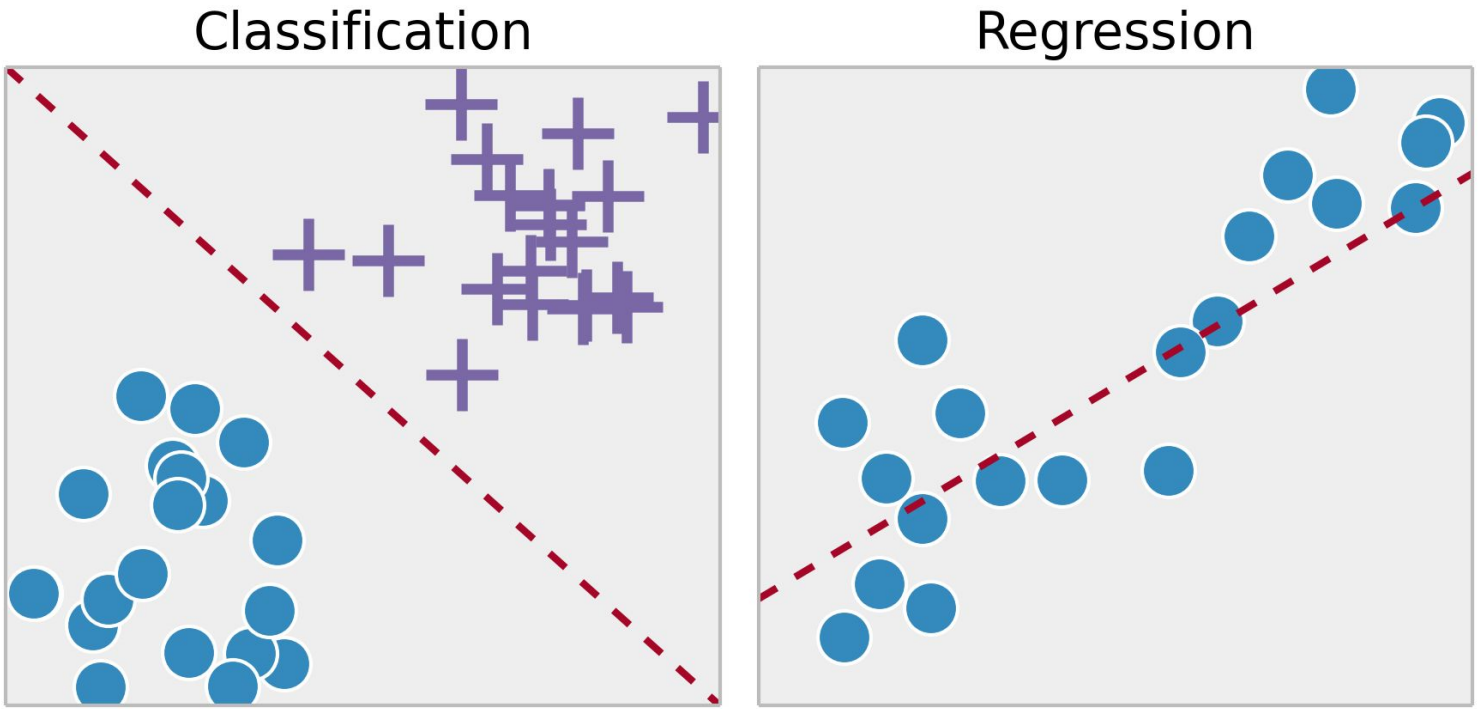
Supervisionado





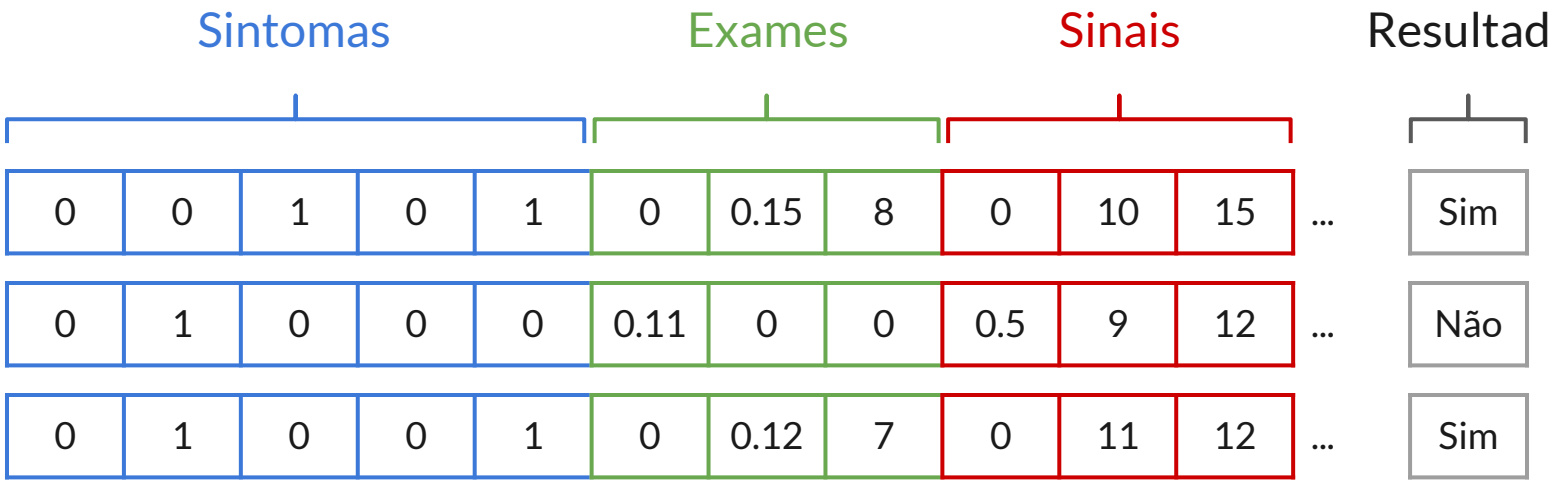
Imagens: https://leonardoaraujosantos.gitbooks.io

Supervisionado

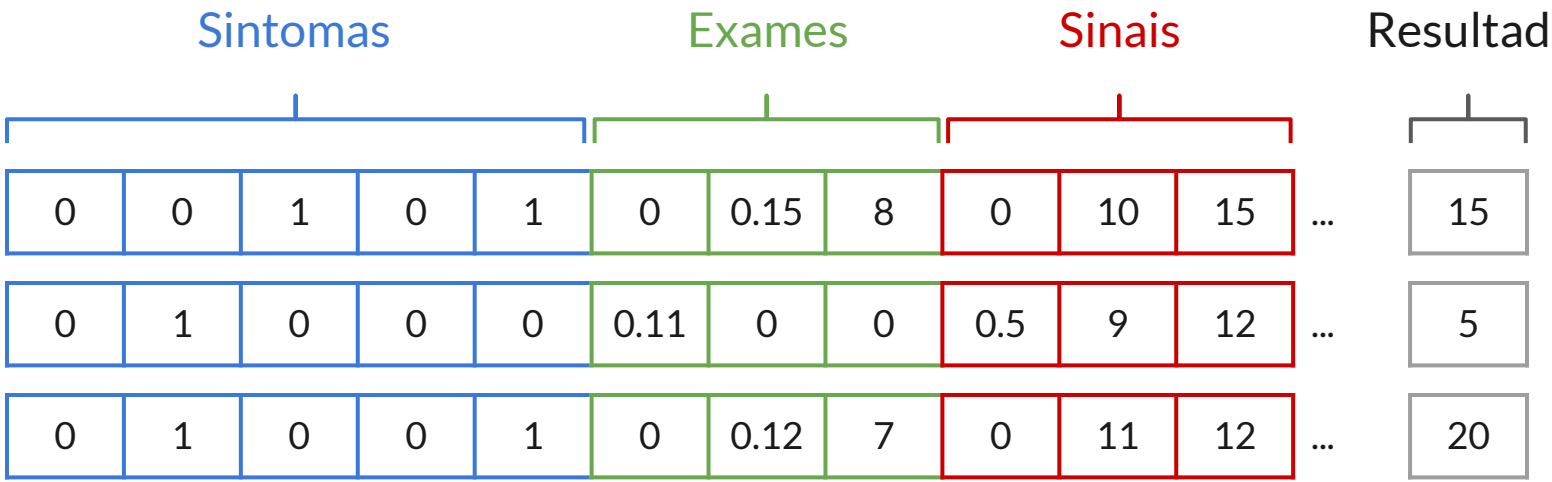


Imagens: https://leonardoaraujosantos.gitbooks.io

Classificação (attributos)

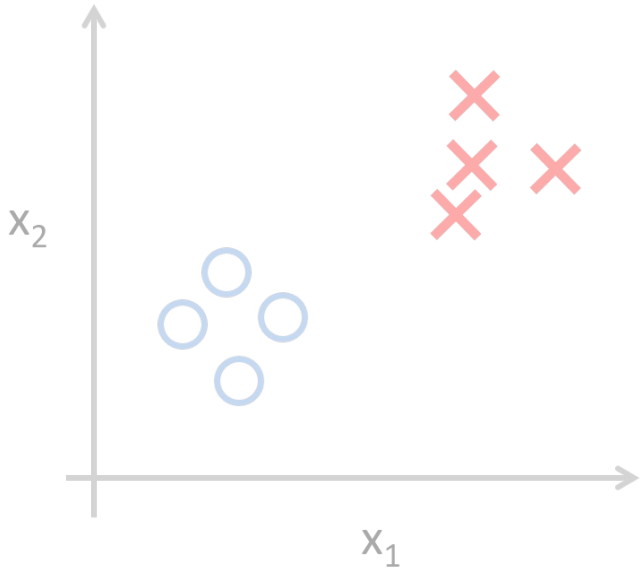


Regressão (attributos)

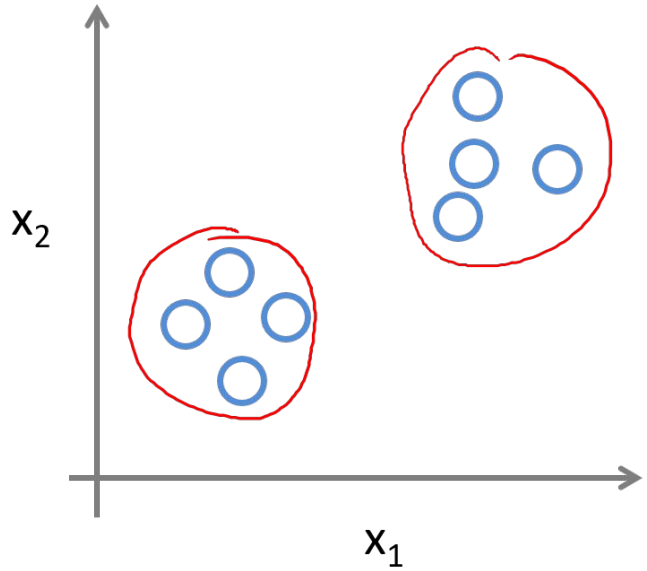


Não-Supervisionado

Supervised Learning

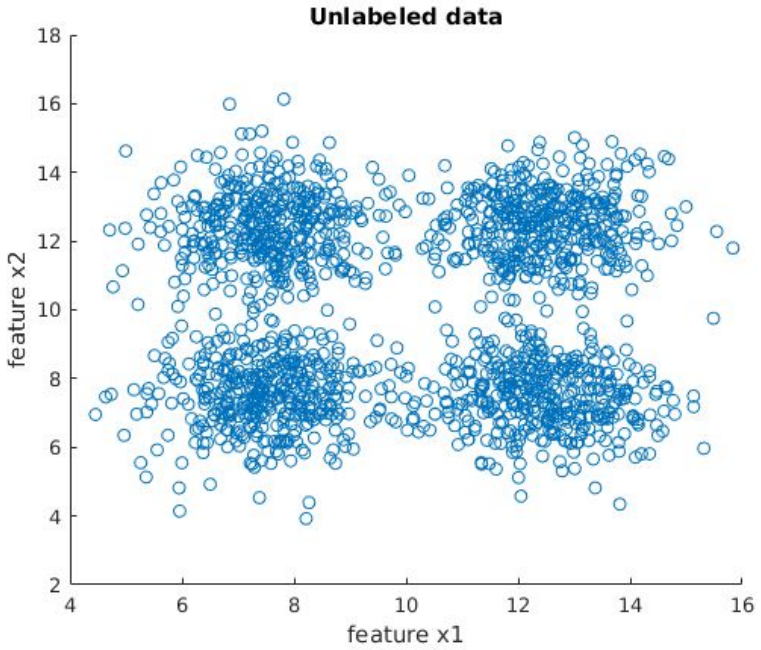


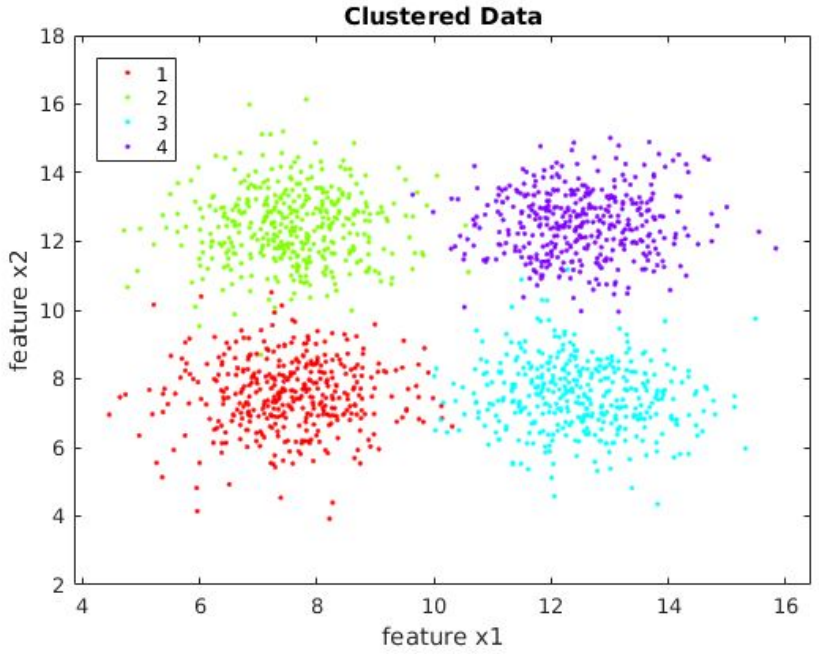
Unsupervised Learning



Imagens: https://leonardoaraujosantos.gitbooks.io

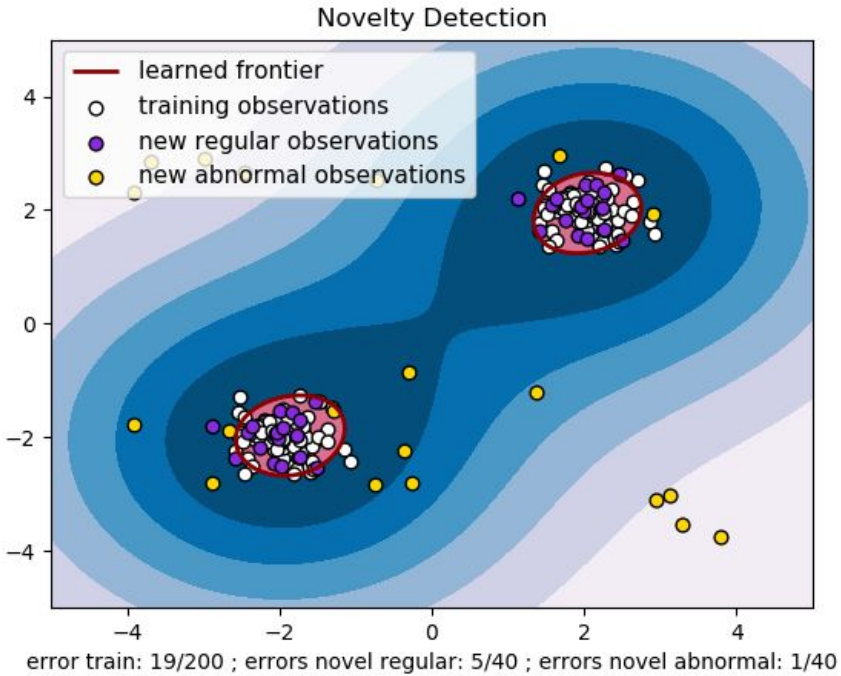
Não-Supervisionado Agrupamento

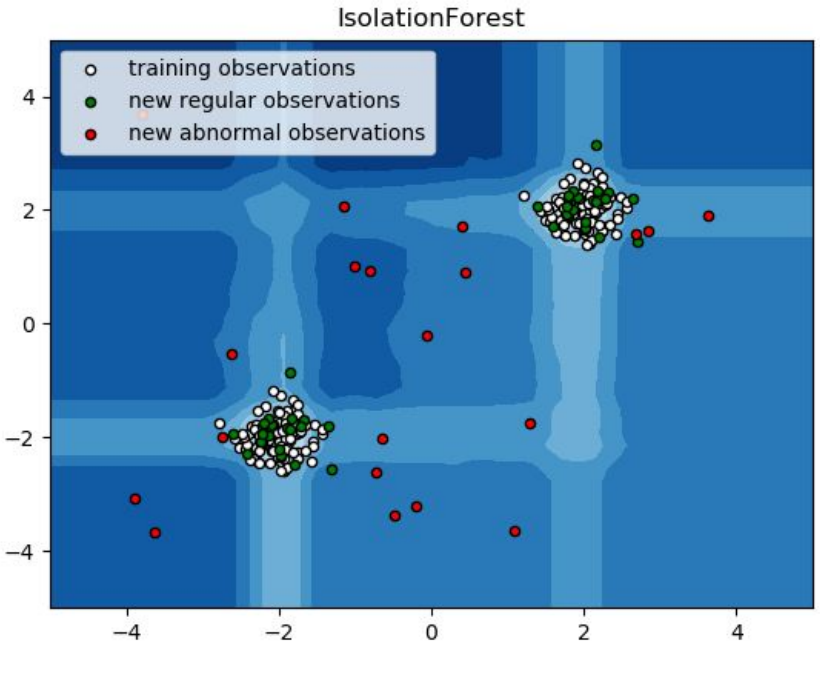




Imagens: https://leonardoaraujosantos.gitbooks.io

Não-Supervisionado Detecção de Anomalia





Imagens: http://scikit-learn.org/stable/modules/outlier_detection.html

Não-Supervisionado Regras de Associação

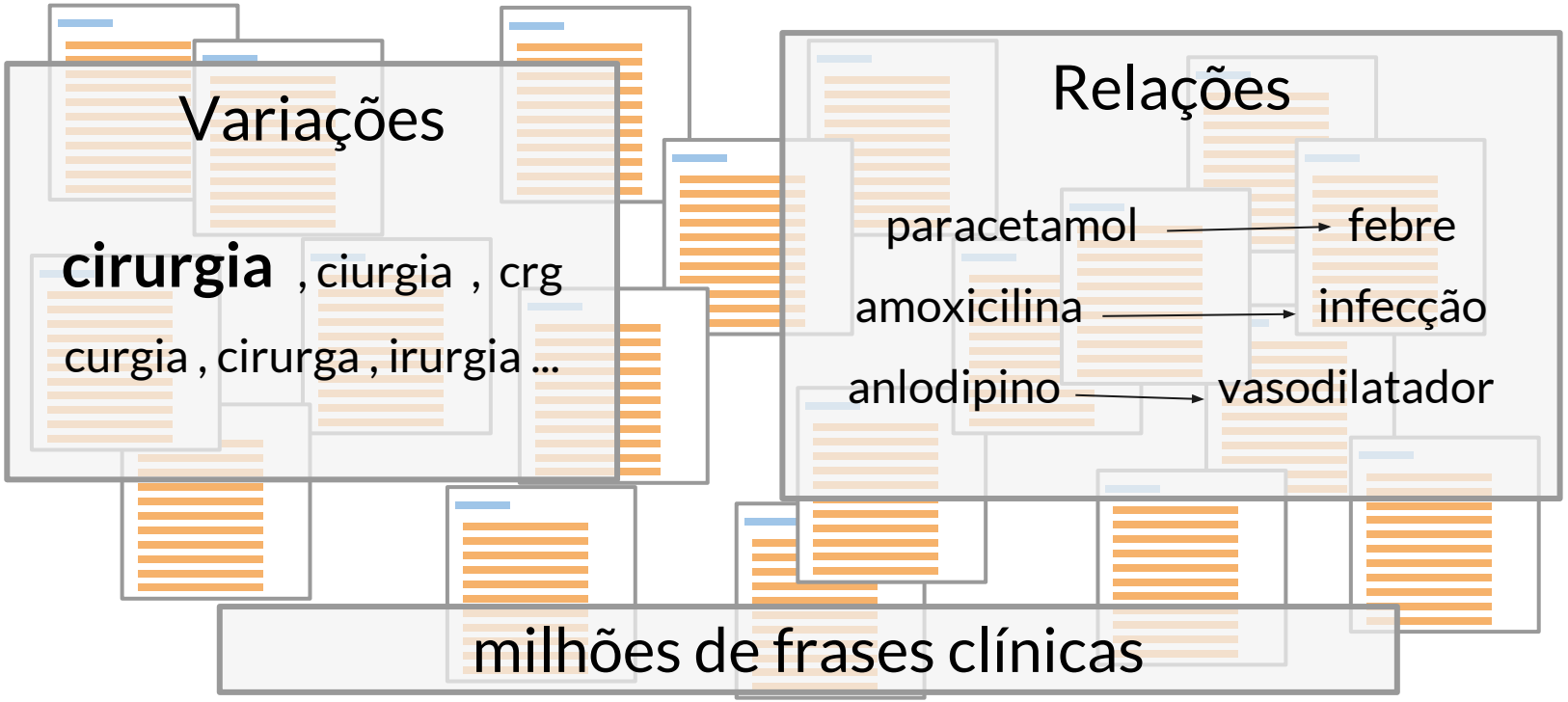
Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B\&C \Rightarrow D$	1/5	1/3	5/9

$$Supprt = \frac{Frequency(X,Y)}{N}$$

$$Confidence = \frac{Frequency(X,Y)}{Frequency(X)}$$

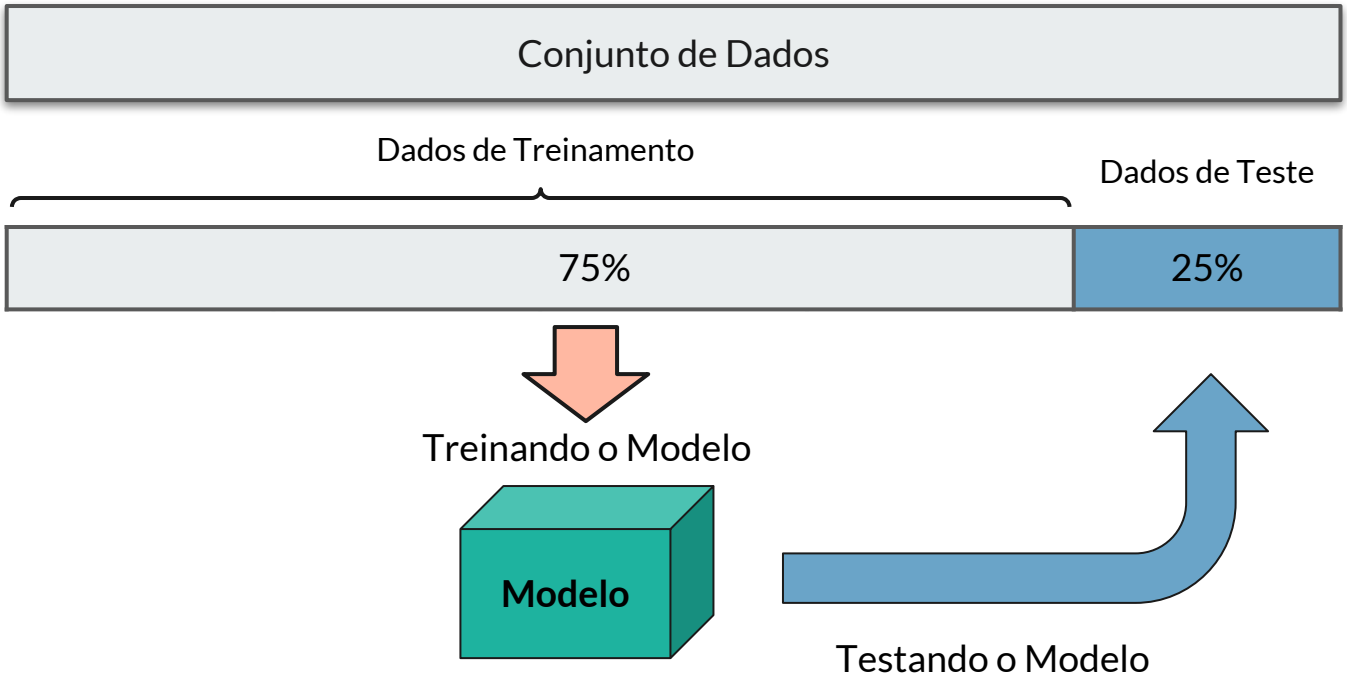
$$Lift = \frac{Support}{Support(X) \times Support(Y)}$$

Não-Supervisionado Word Embeddings

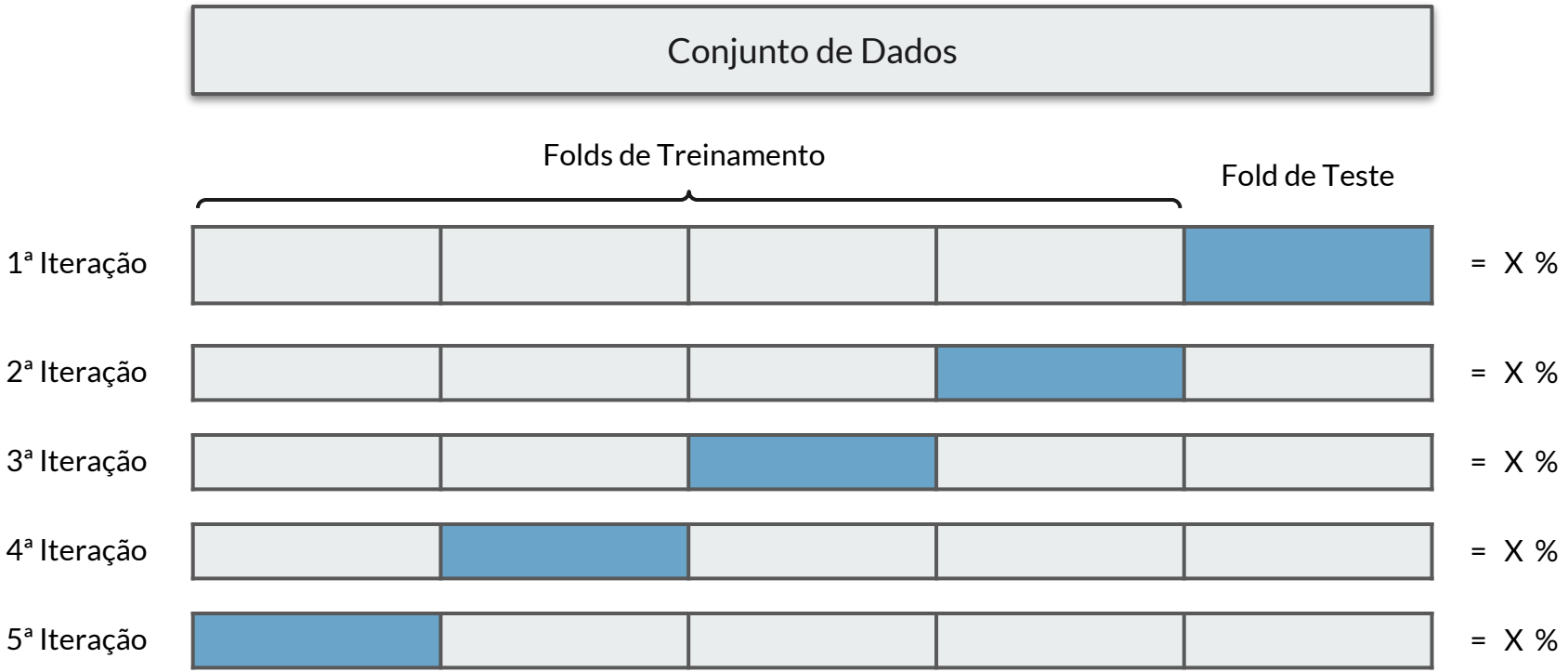


Avaliação dos Modelos

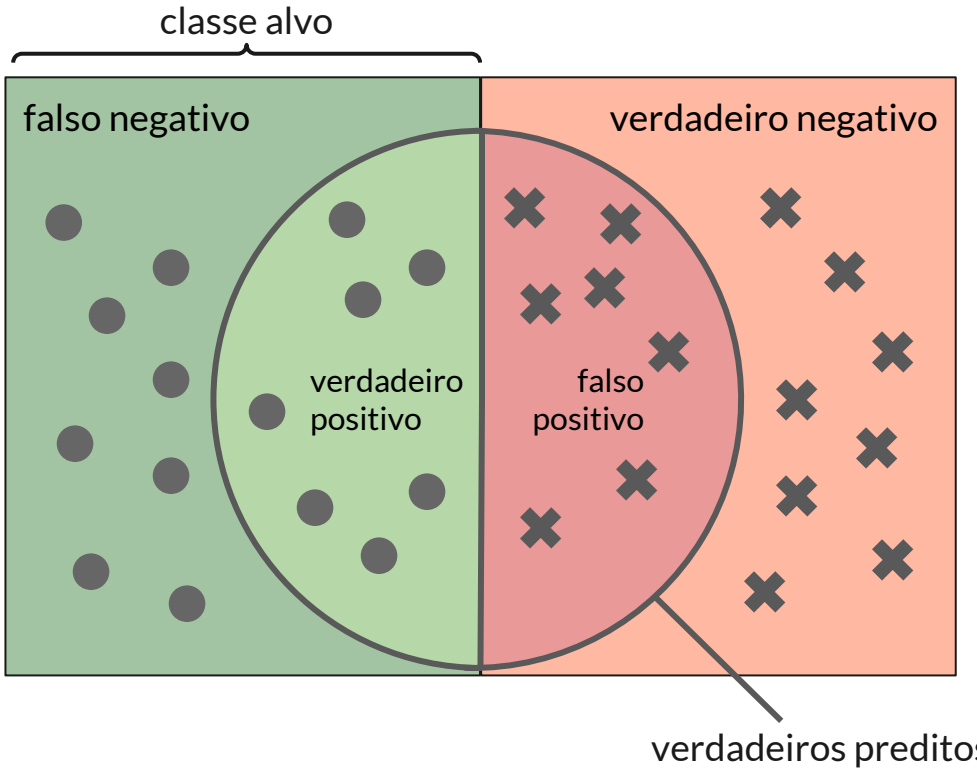
Treino e Teste



Validação Cruzada (5 Folds)



Acurácia, Precisão e Abrangência Classificação

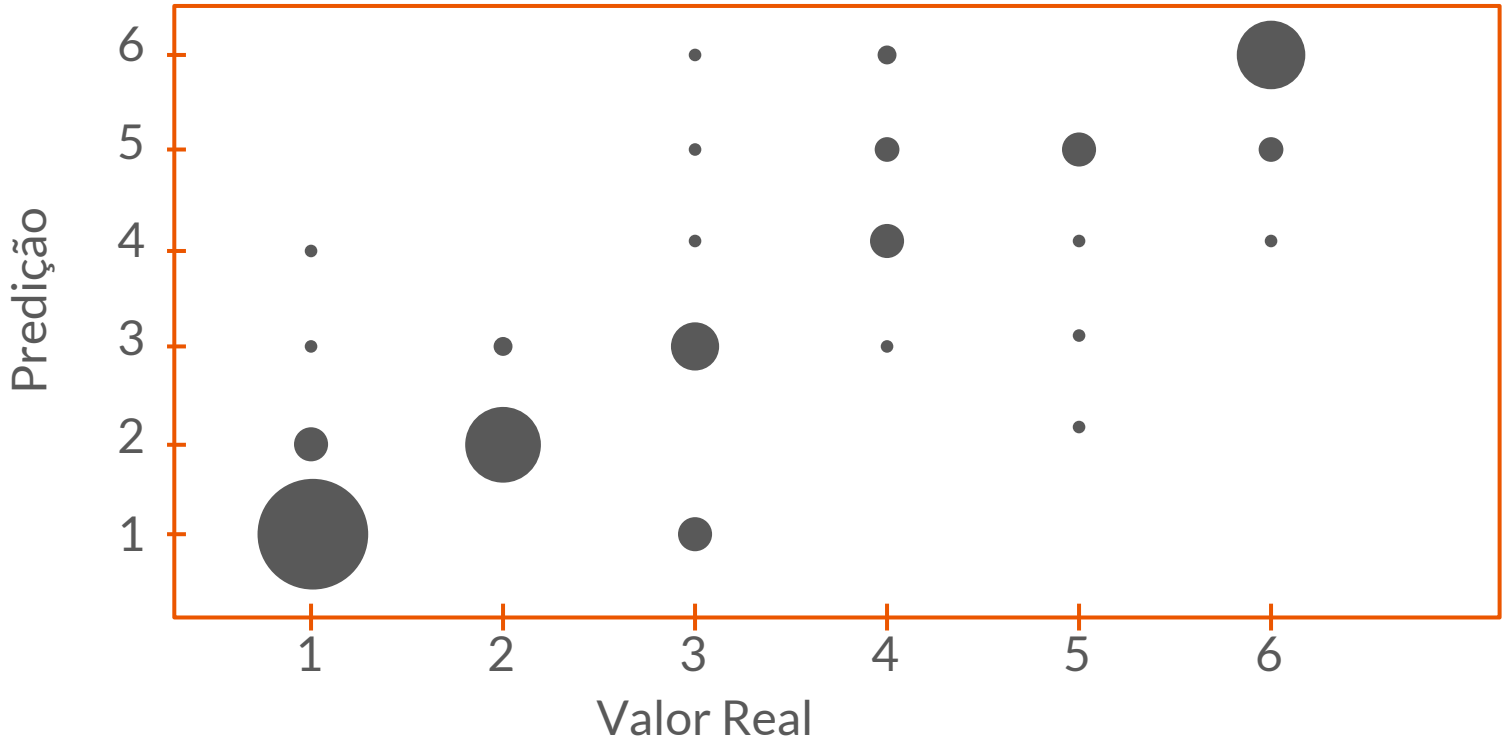


Acurácia = $\frac{\text{verdadeiro positivo} + \text{verdadeiro negativo}}{\text{todos}}$

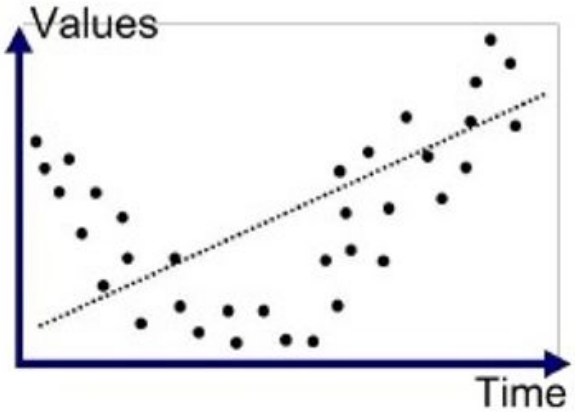
Precisão = $\frac{\text{verdadeiro positivo}}{\text{verdadeiro positivo} + \text{falso positivo}}$

Abrangência = $\frac{\text{verdadeiro positivo}}{\text{verdadeiro positivo} + \text{falso negativo}}$

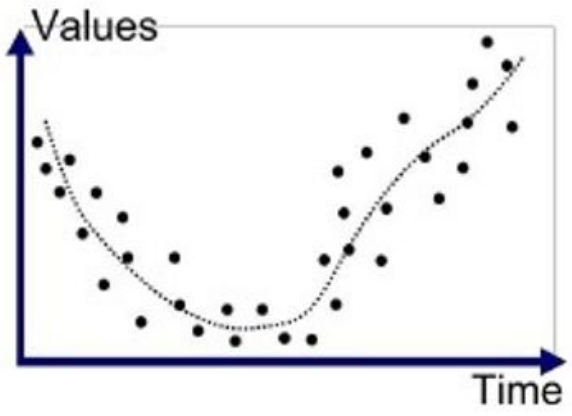
Erro Médio Absoluto Regressão



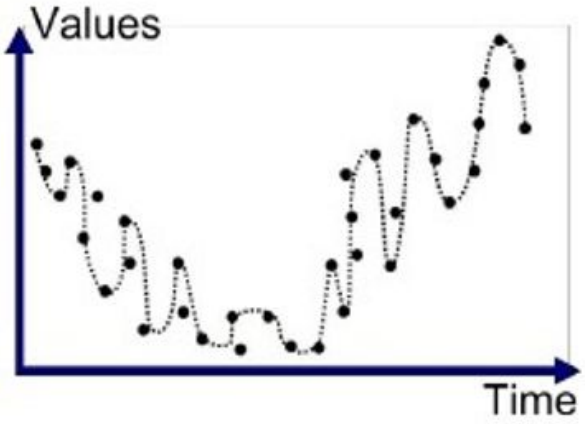
Overfitting na Regressão



Underfitted

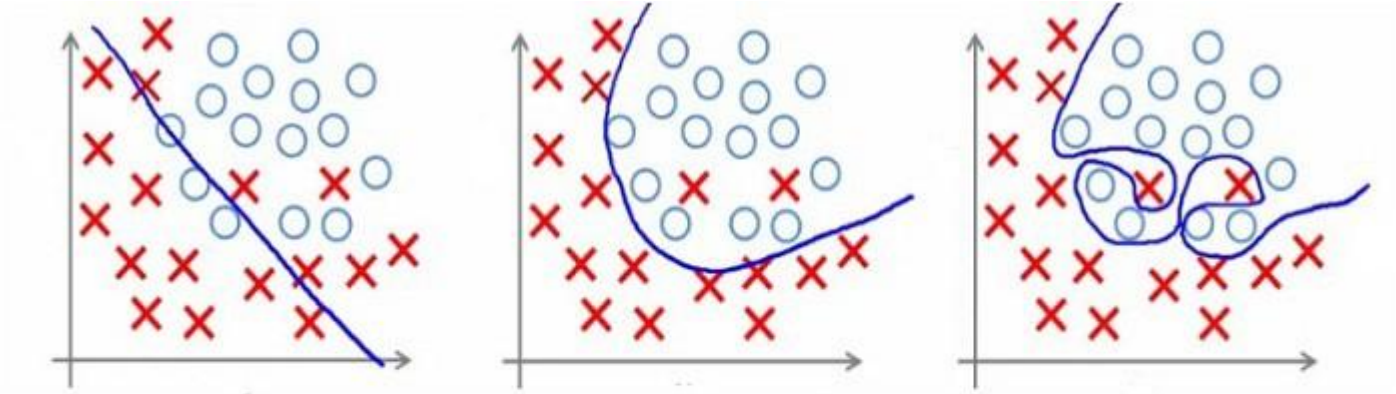


Good Fit/Robust



Overfitted

Overfitting na Classificação



Under-fitting

(too simple to explain the variance)

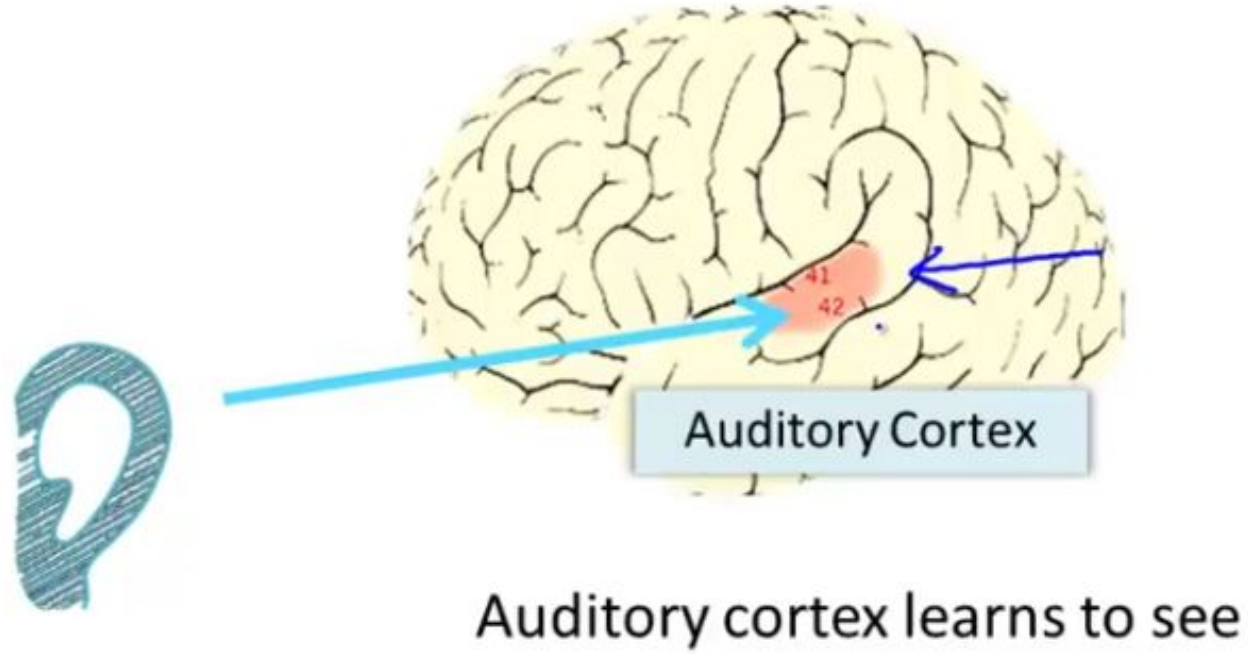
Appropriate-fitting

Over-fitting

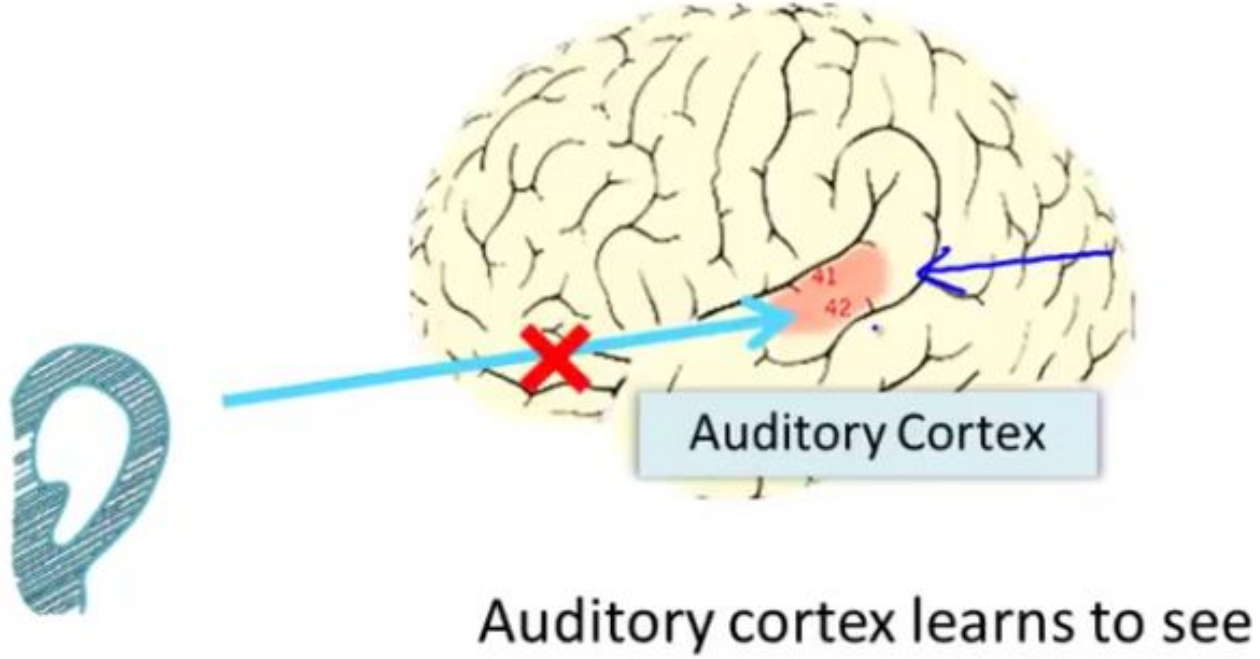
(forcefitting -- too good to be true)

Redes Neurais

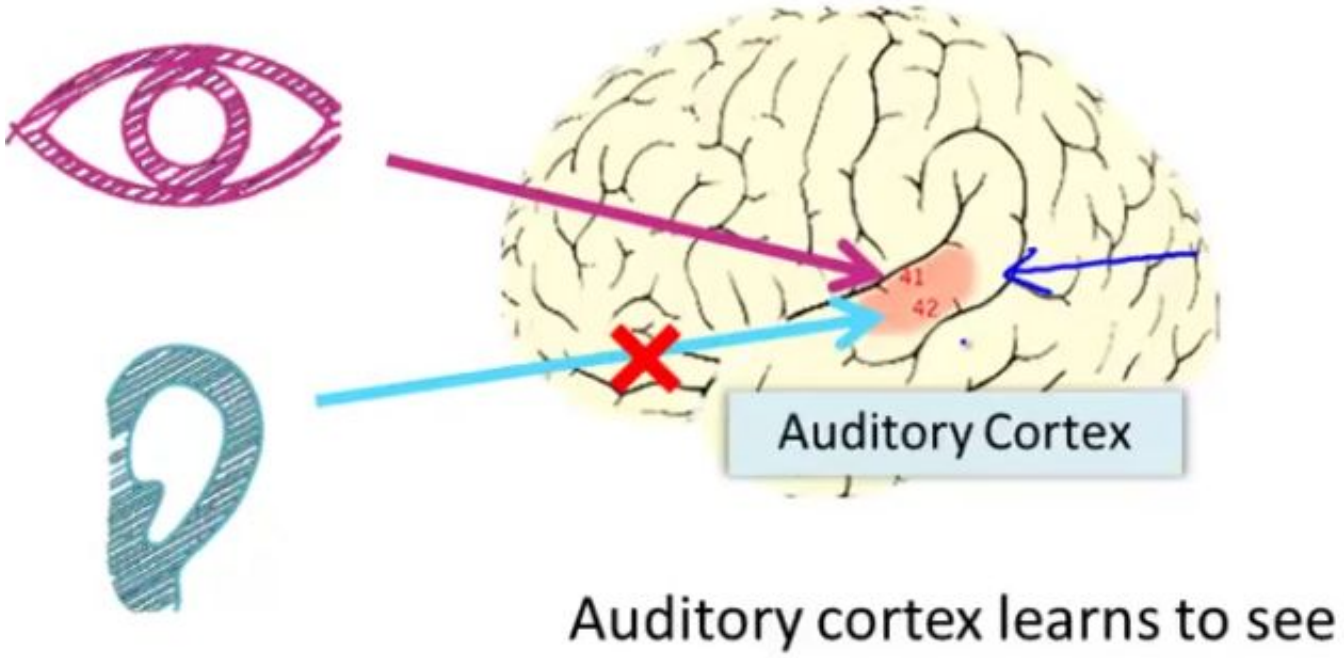
A hipótese do algoritmo de aprendizado único



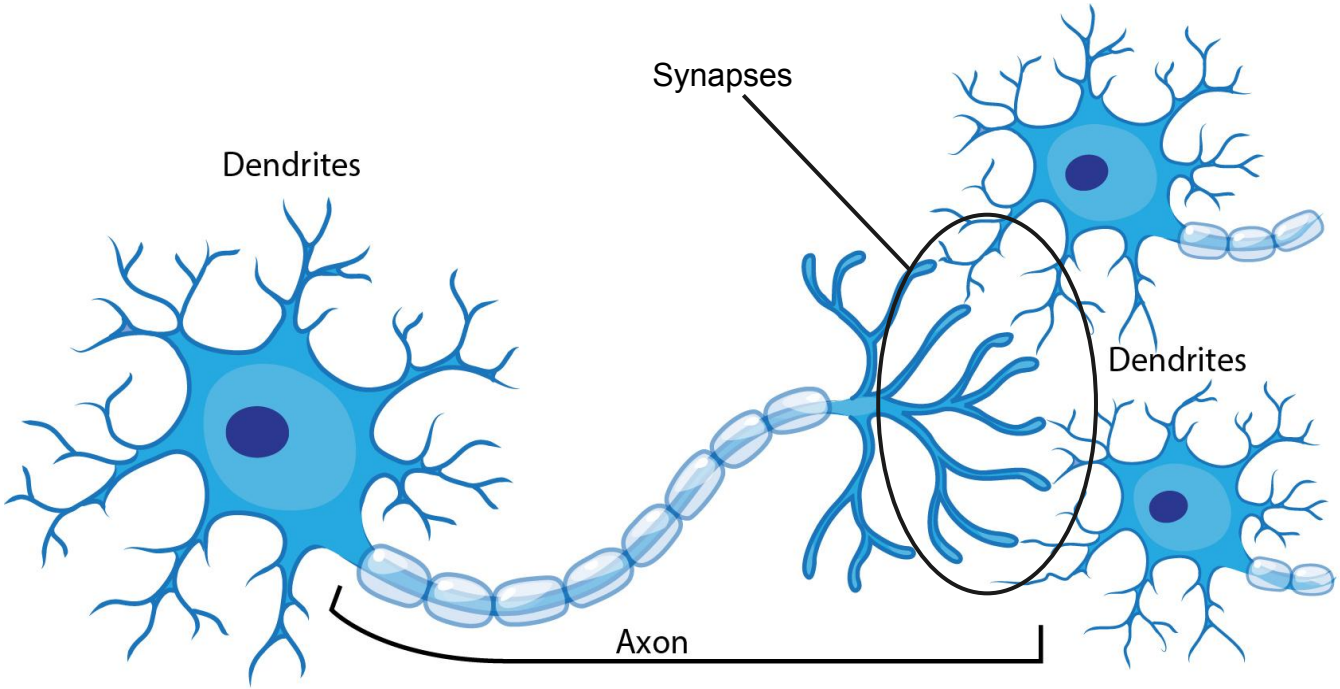
A hipótese do algoritmo de aprendizado único



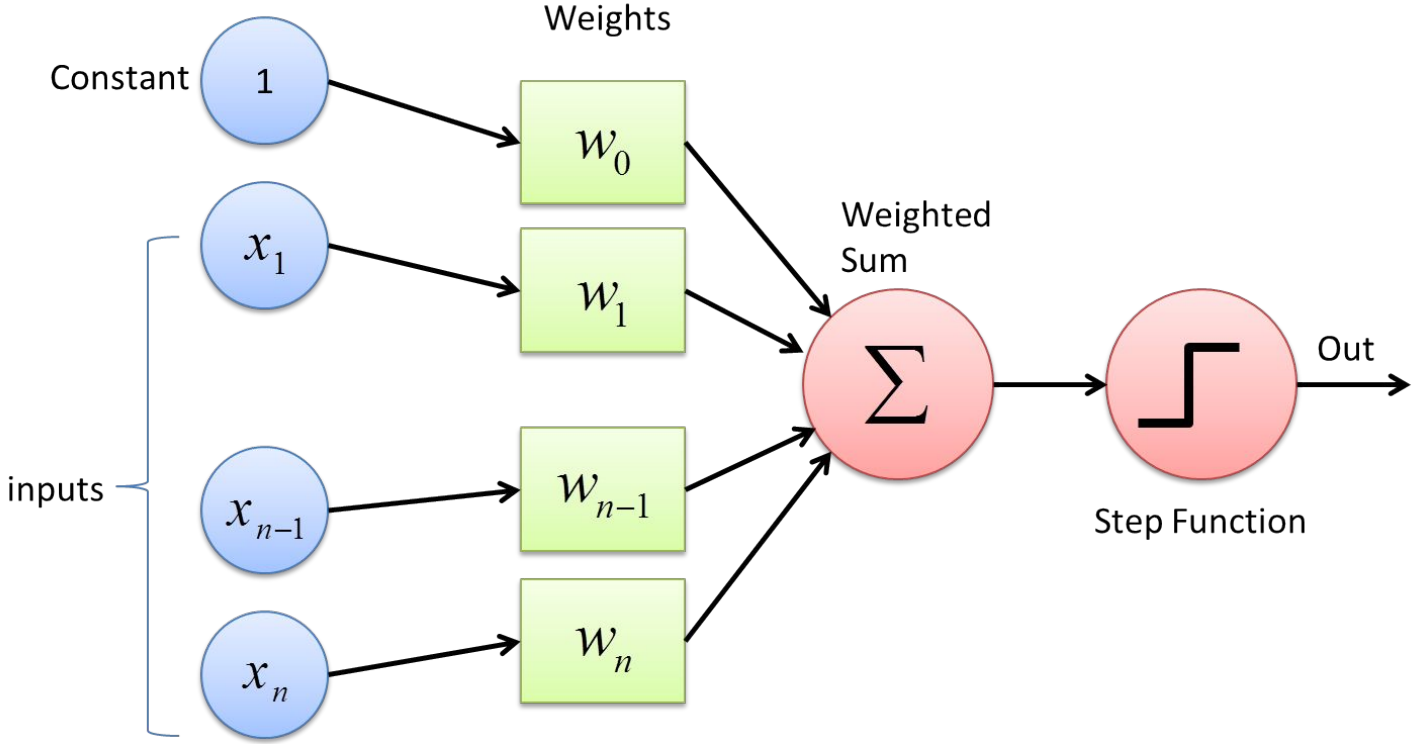
A hipótese do algoritmo de aprendizado único



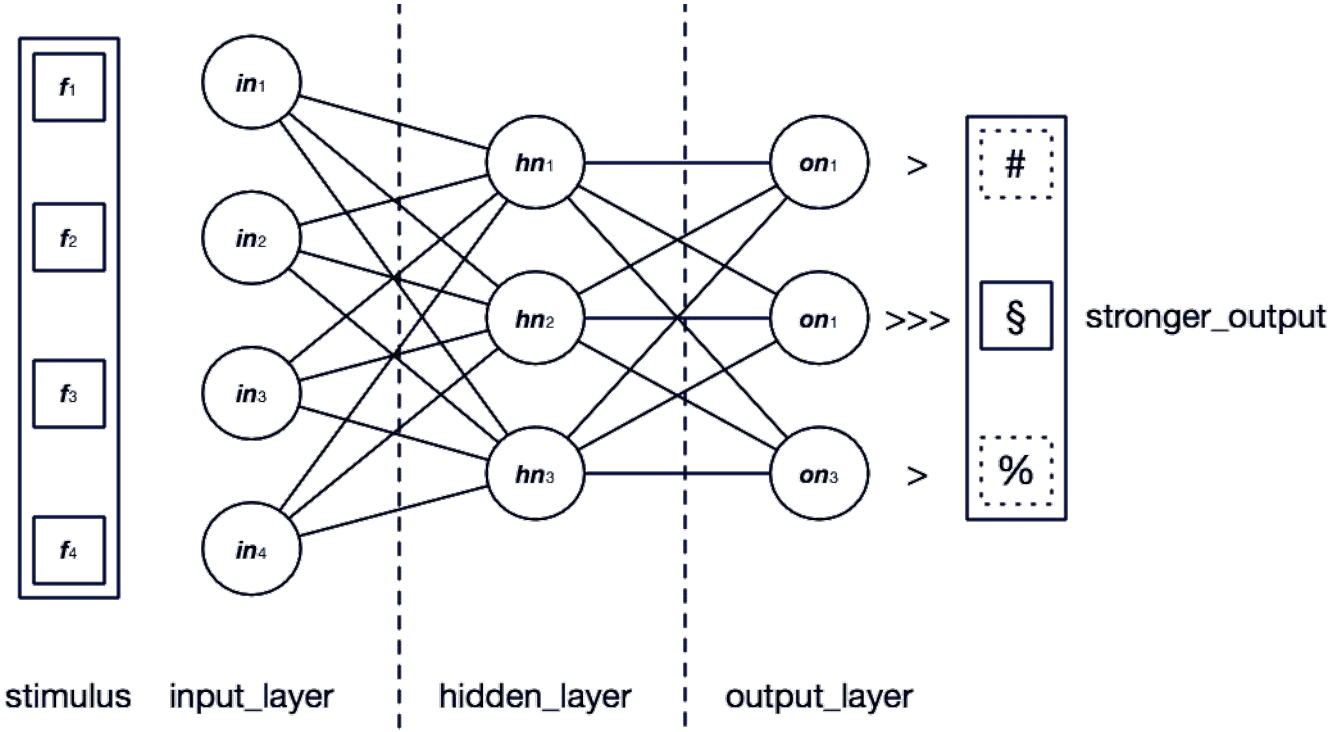
Neurônio (perceptron)



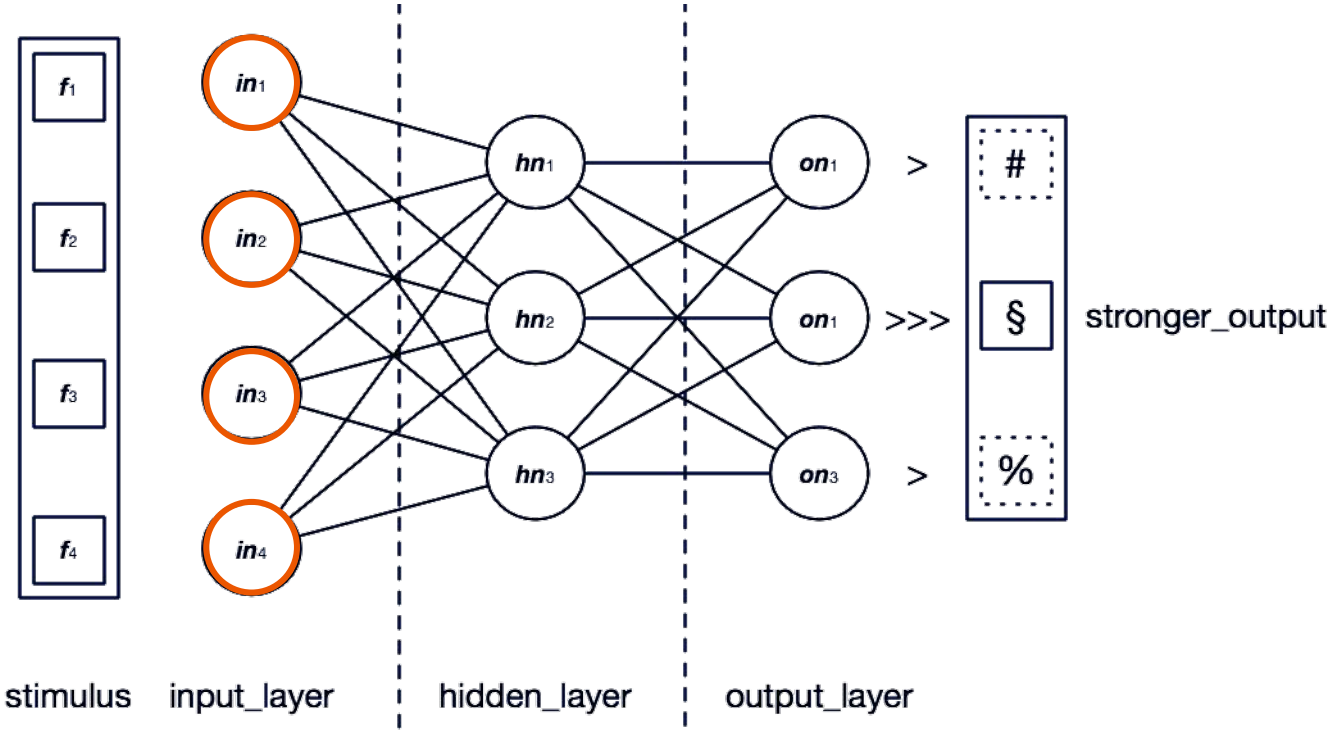
Perceptron (neurônio)



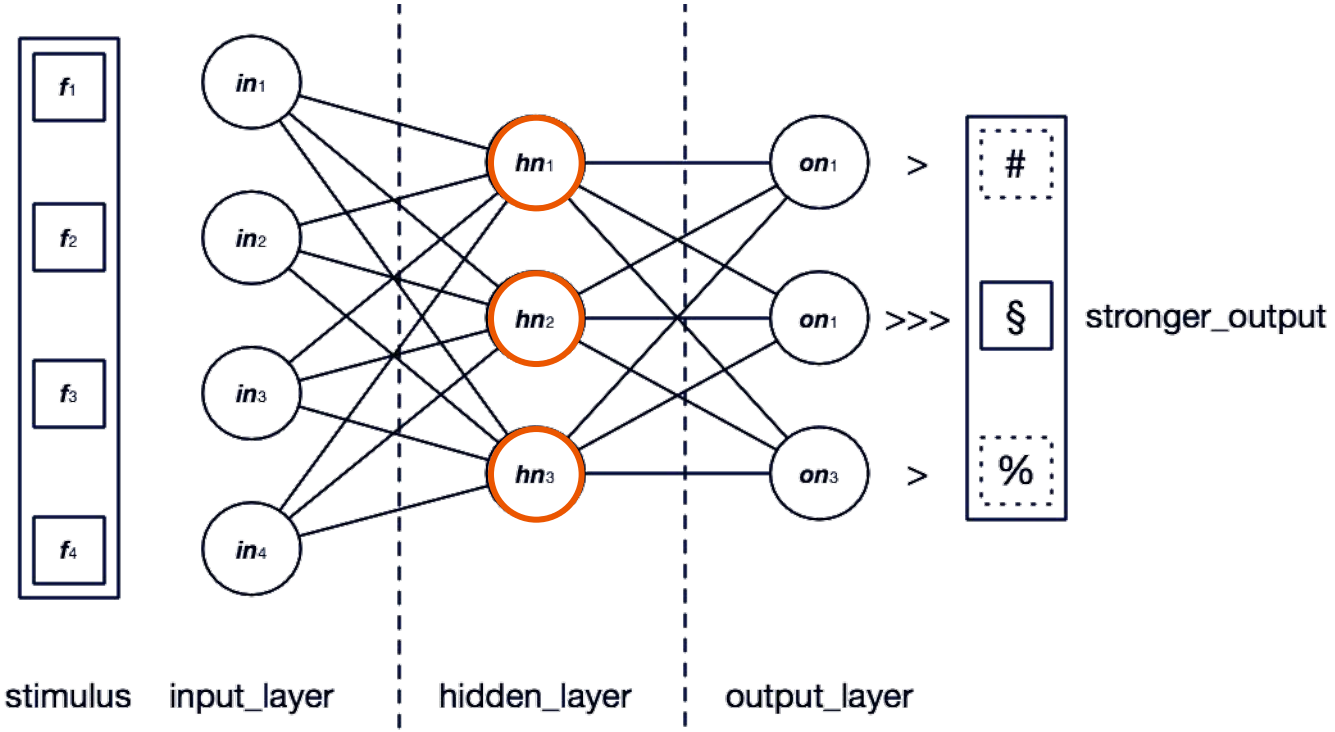
Multi Layer Perceptron



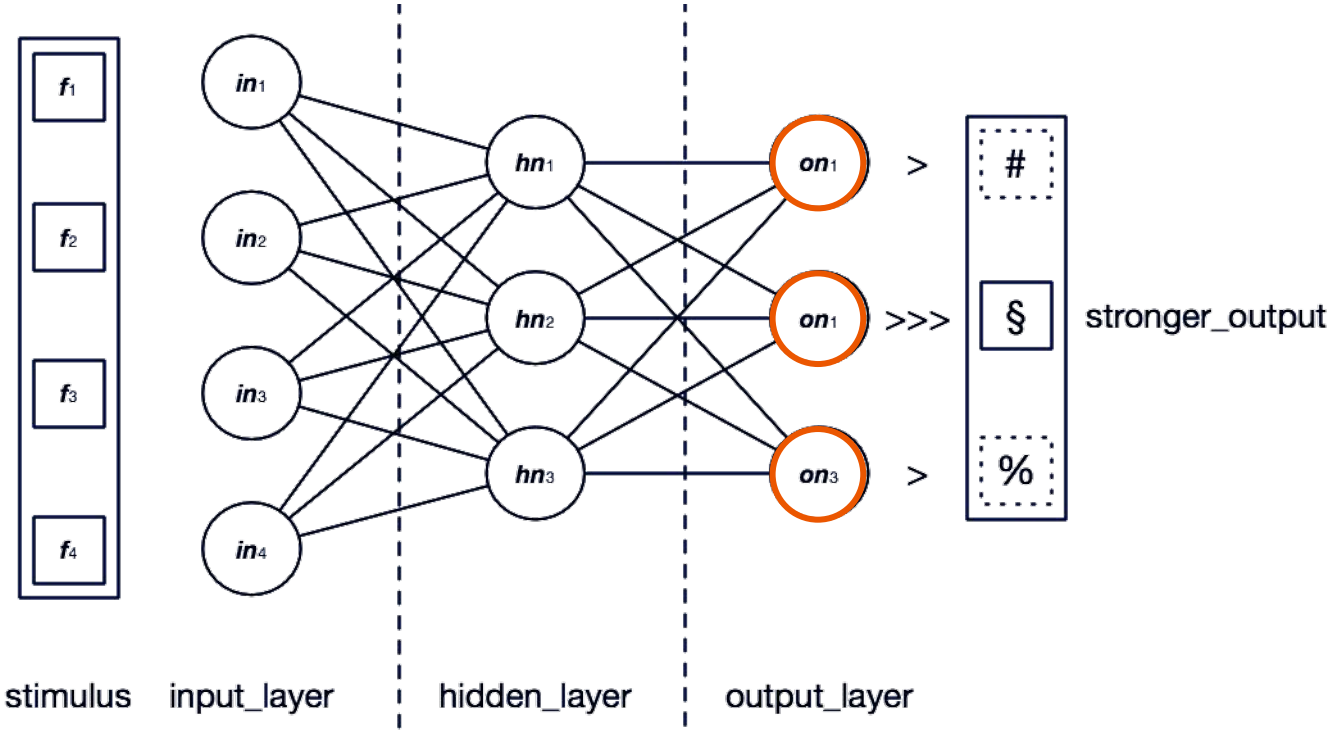
Multi Layer Perceptron



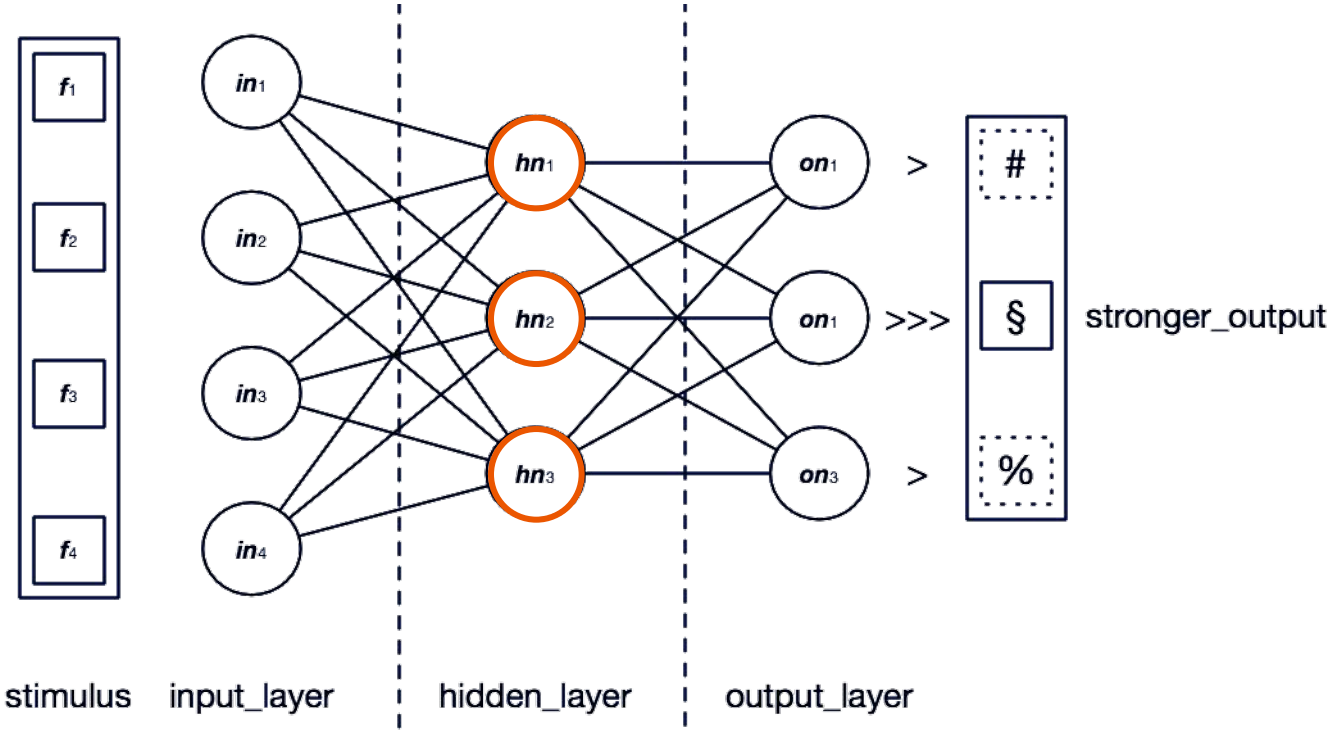
Multi Layer Perceptron



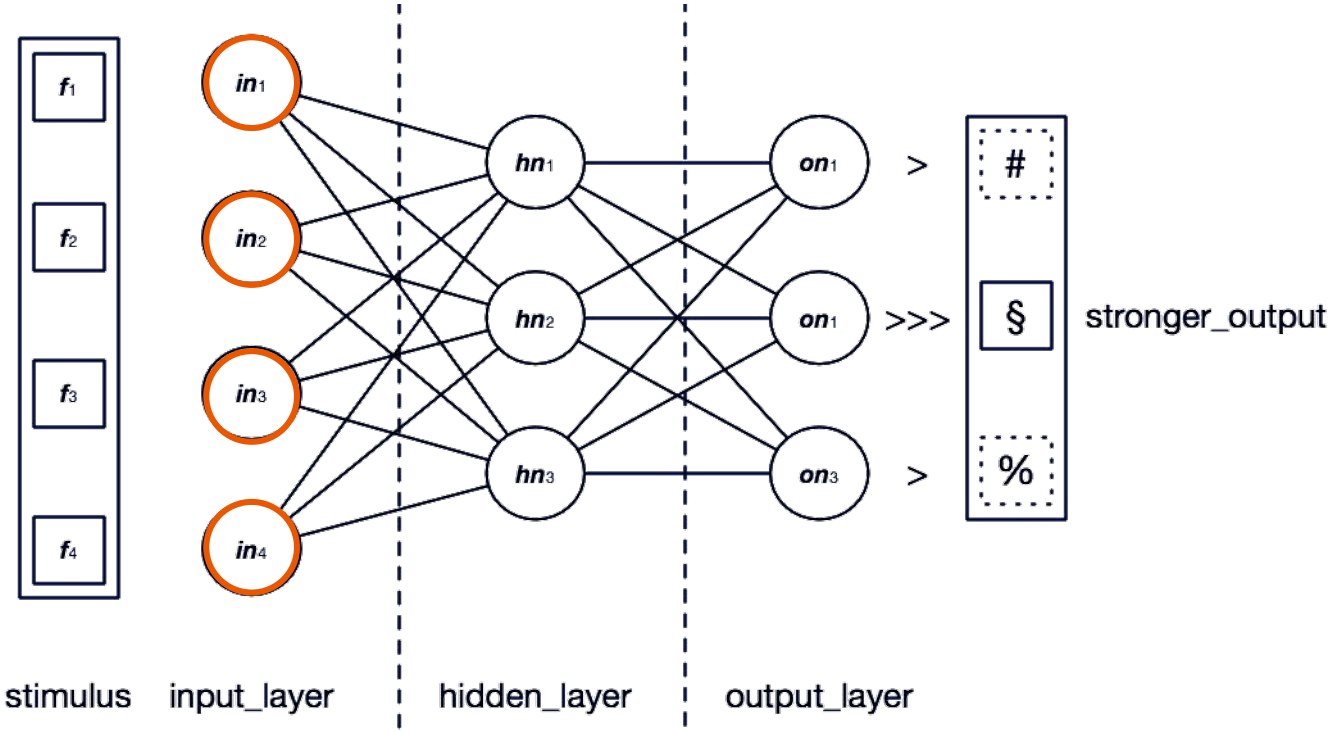
Multi Layer Perceptron



Multi Layer Perceptron



Multi Layer Perceptron



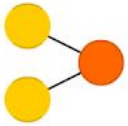
Do que as RNs são capazes?

- Why Deep Learning Now? - ColdFusion

Neural Networks

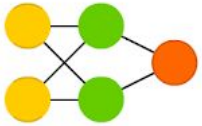
- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

Perceptron (P)

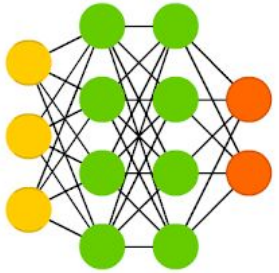


Feed Forward (FF)

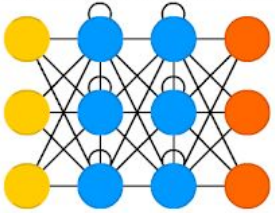
Radial Basis Network (RBF)



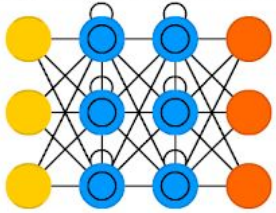
Deep Feed Forward (DFF)



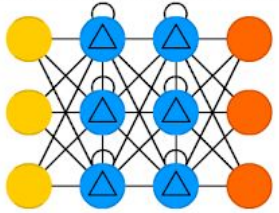
Recurrent Neural Network (RNN)



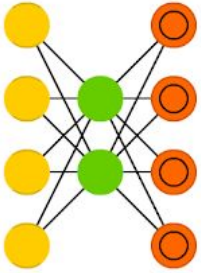
Long / Short Term Memory (LSTM)



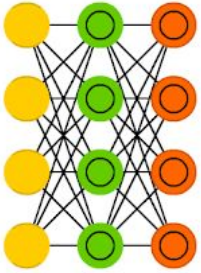
Gated Recurrent Unit (GRU)



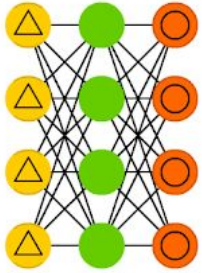
Auto Encoder (AE)



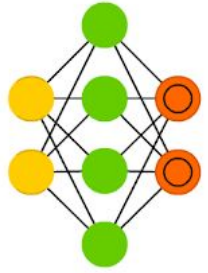
Variational AE (VAE)



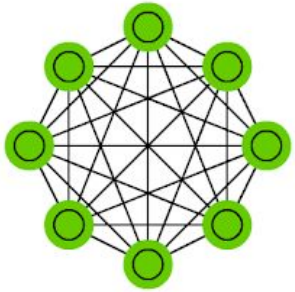
Denoising AE (DAE)



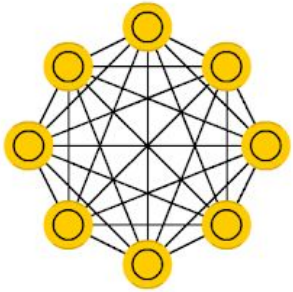
Sparse AE (SAE)



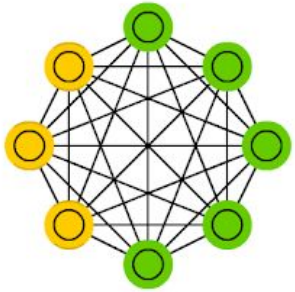
Markov Chain (MC)



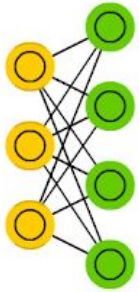
Hopfield Network (HN)



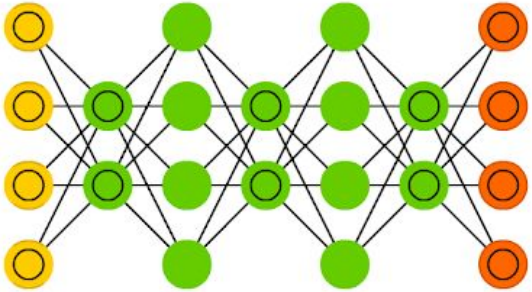
Boltzmann Machine (BM)



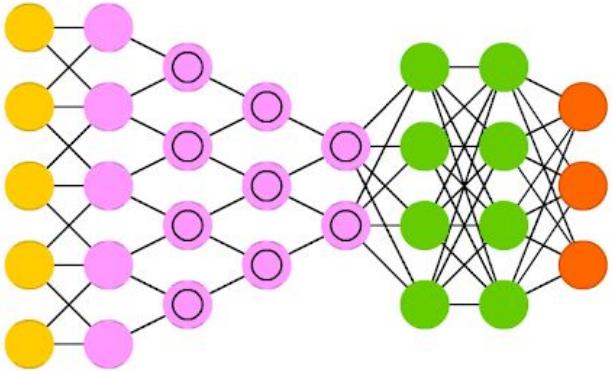
Restricted BM (RBM)



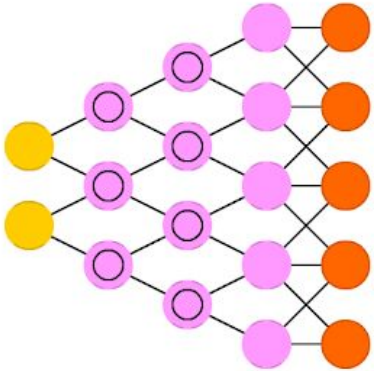
Deep Belief Network (DBN)



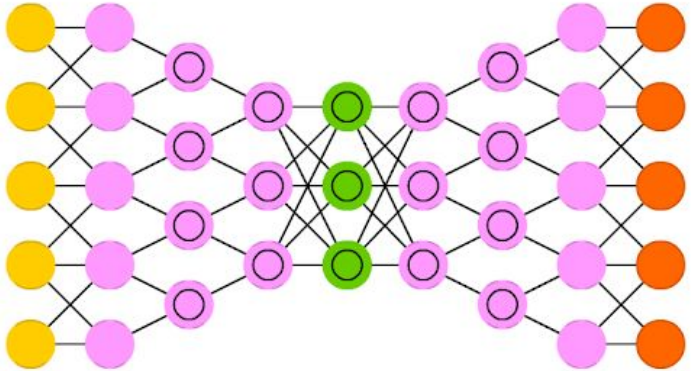
Deep Convolutional Network (DCN)



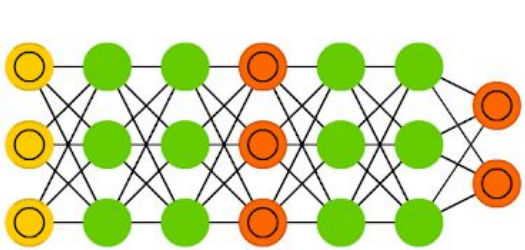
Deconvolutional Network (DN)



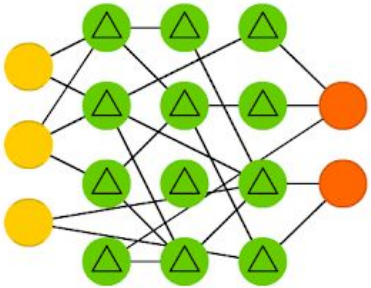
Deep Convolutional Inverse Graphics Network (DCIGN)



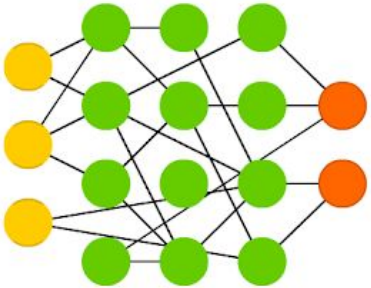
Generative Adversarial Network (GAN)



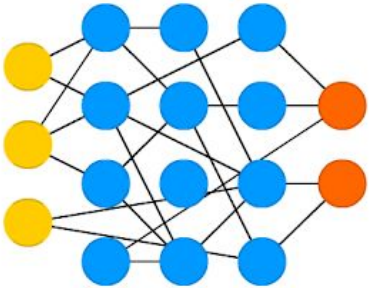
Liquid State Machine (LSM)



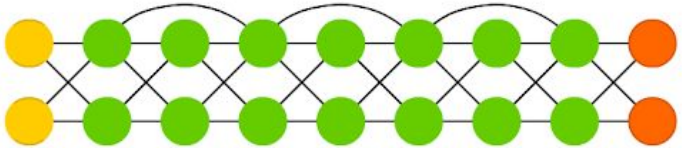
Extreme Learning Machine (ELM)



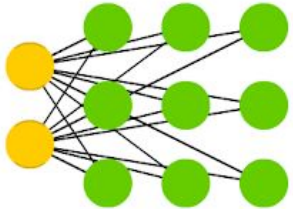
Echo State Network (ESN)



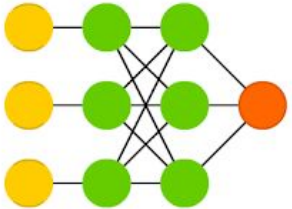
Deep Residual Network (DRN)



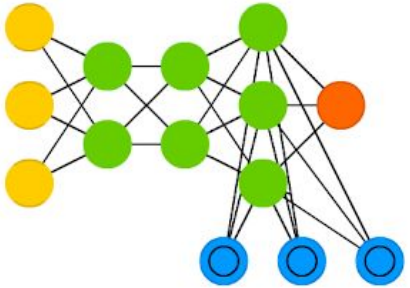
Kohonen Network (KN)



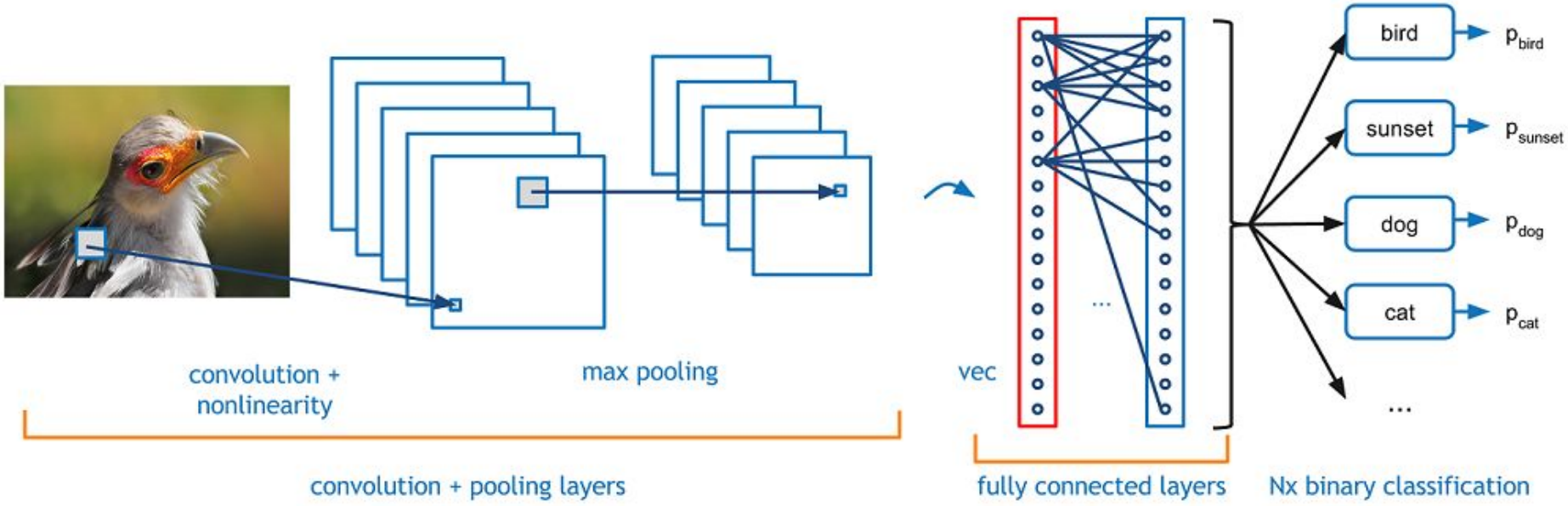
Support Vector Machine (SVM)



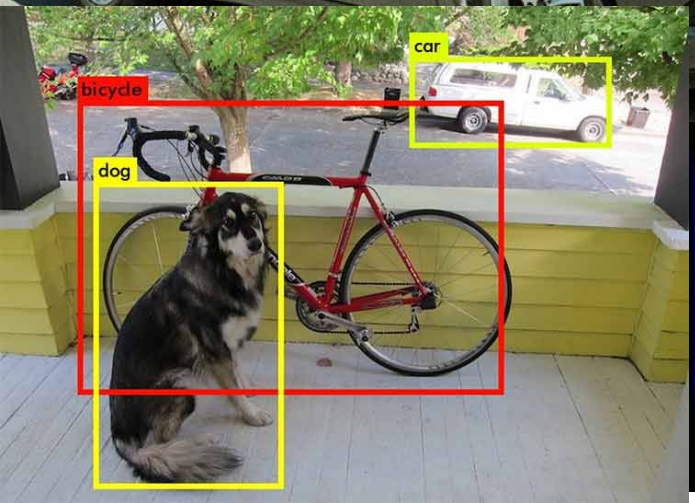
Neural Turing Machine (NTM)

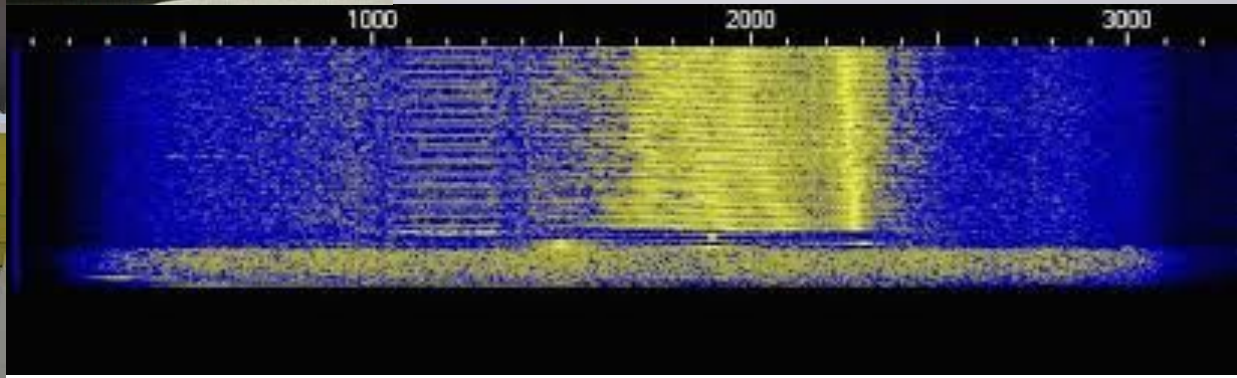


Redes Neurais Convolucionais

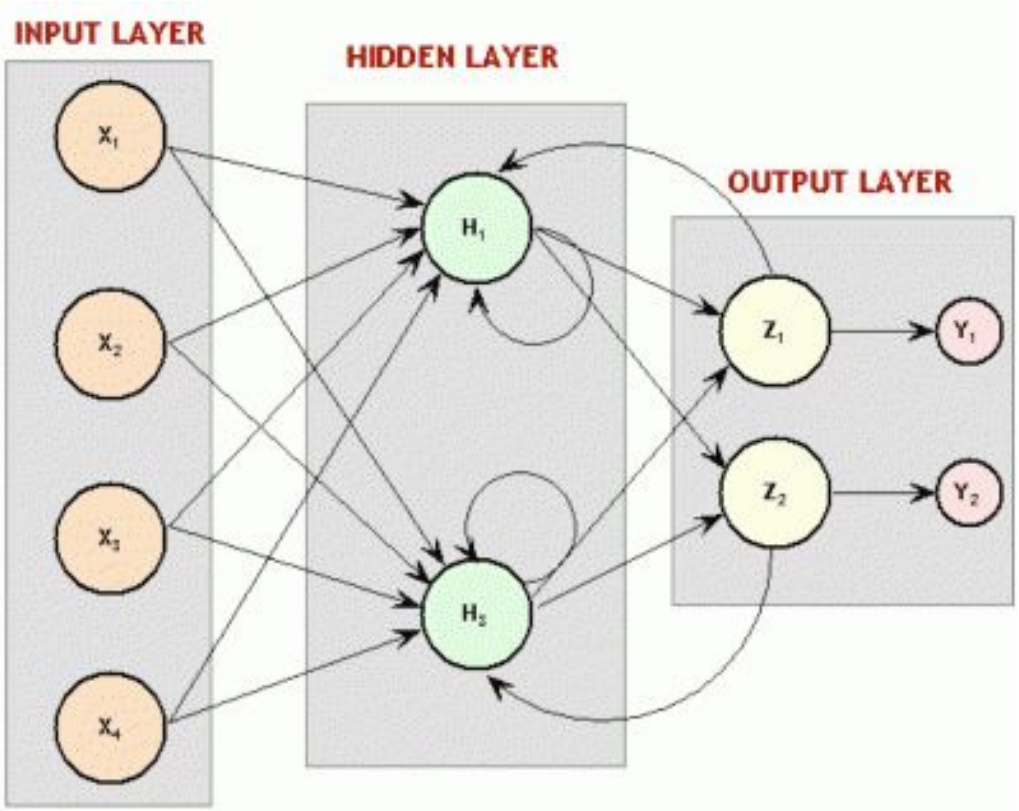



UBER
UBERATC.COM/CAR
ADVANCED
TECHNOLOGIES
CENTER
UBER

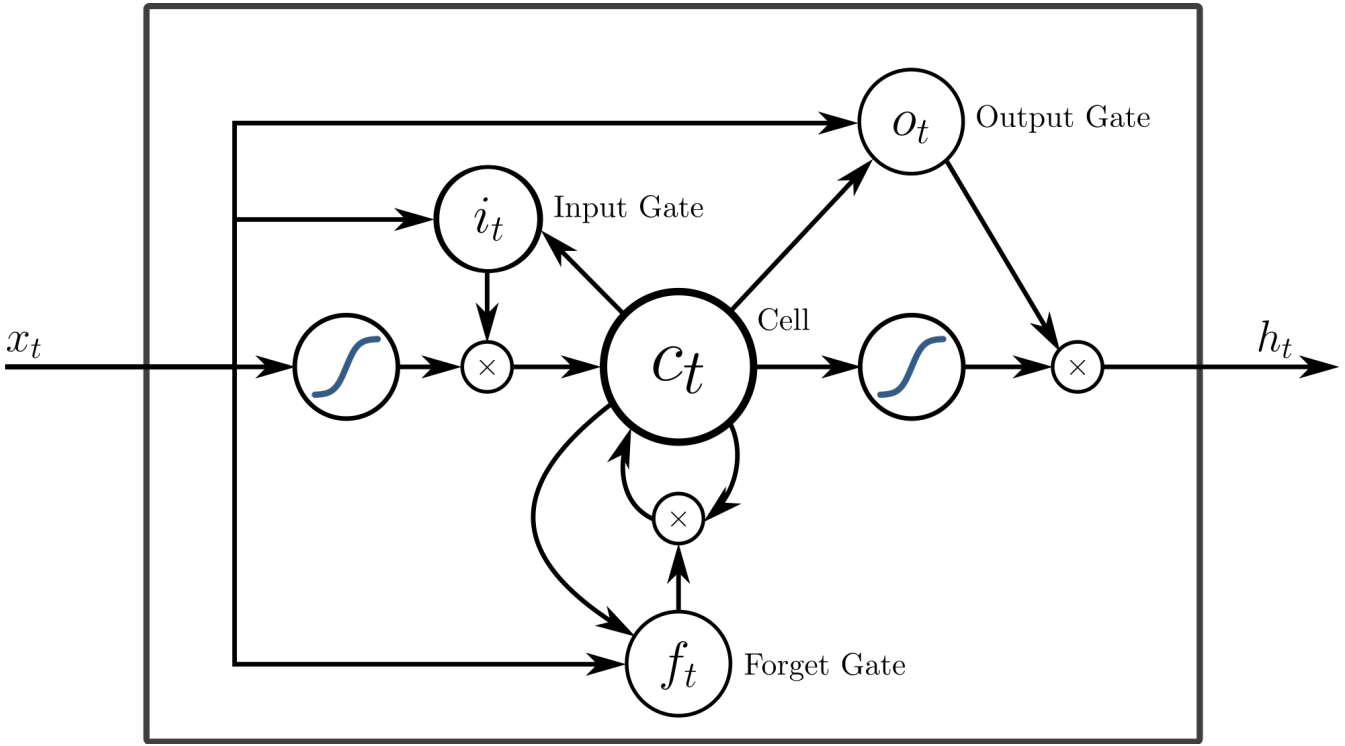

car
bicycle
dog


1000
2000
3000

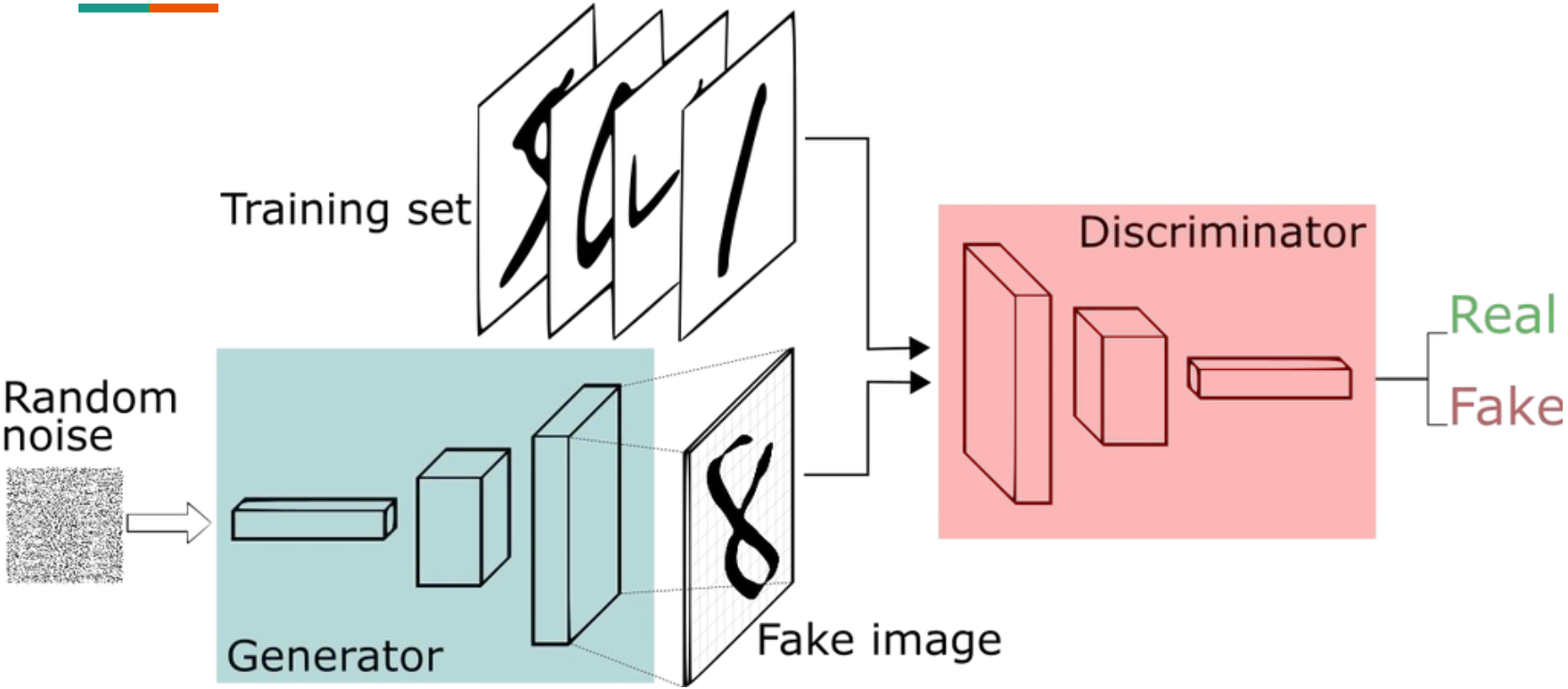
Redes Neurais Recorrentes

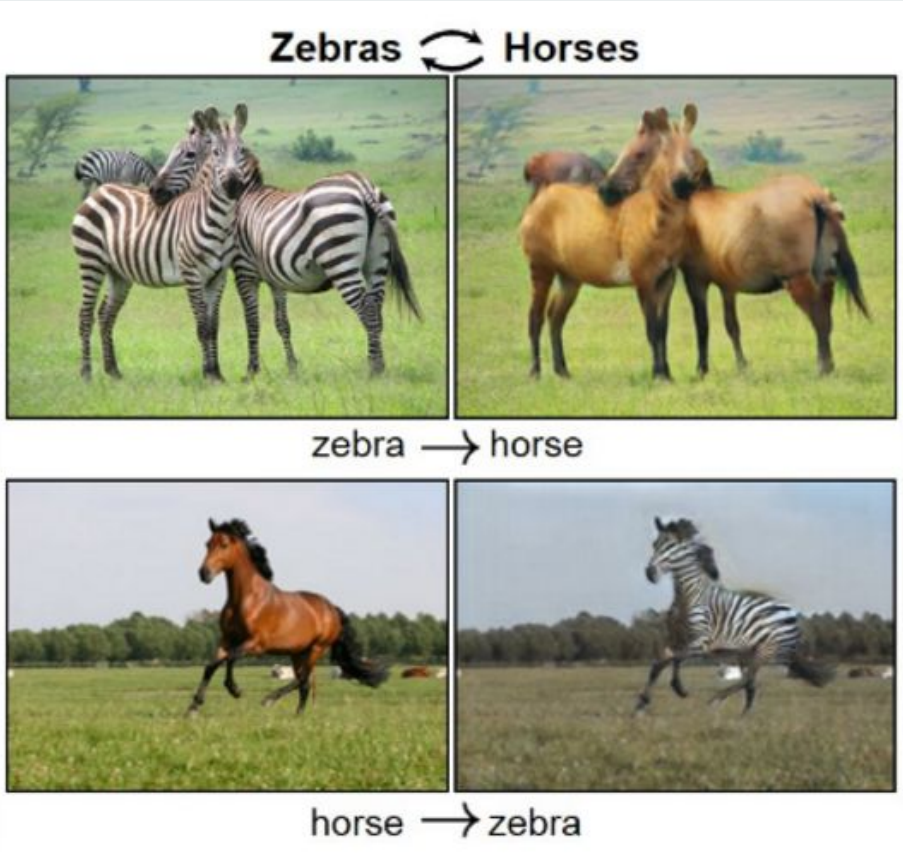


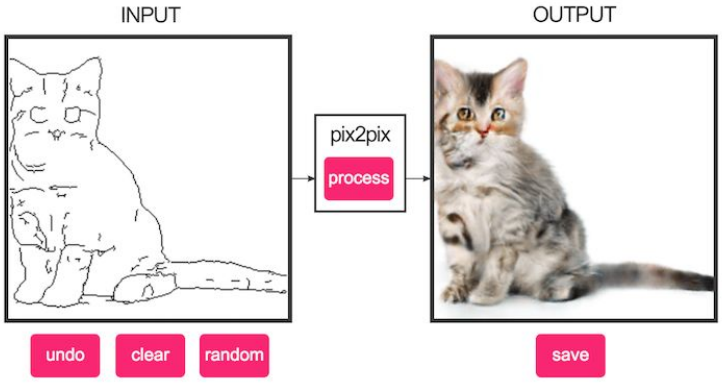
Long-Short Term Memory (LSTM)



Generative Adversarial Network




Zebras ⇄ Horses
zebra → horse
horse → zebra


INPUT
OUTPUT
pix2pix
process
undo
clear
random
save

Testando Redes Neurais

https://playground.tensorflow.org

Limitações das NNs

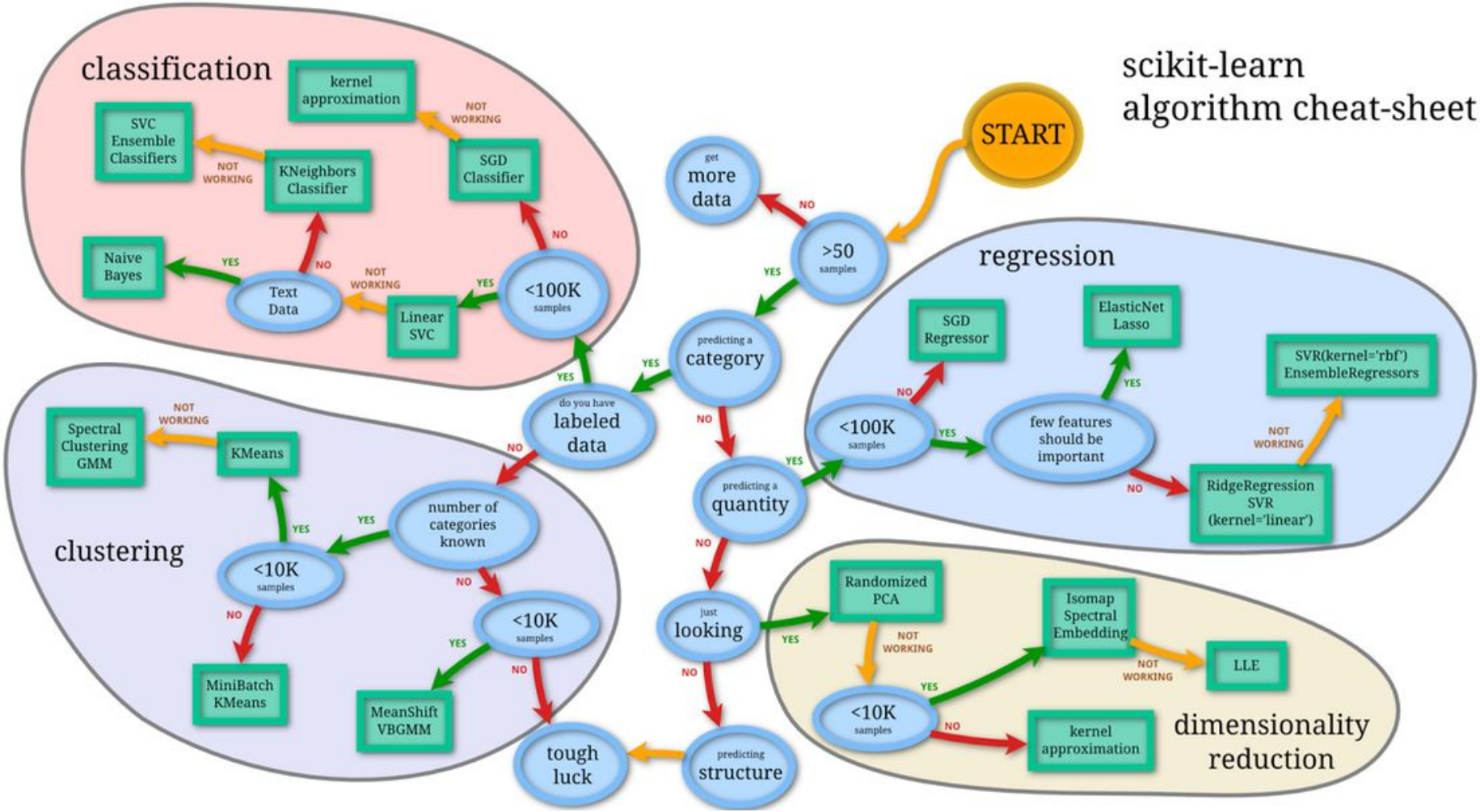
Obrigado!

camila.kolling@acad.pucrs.br

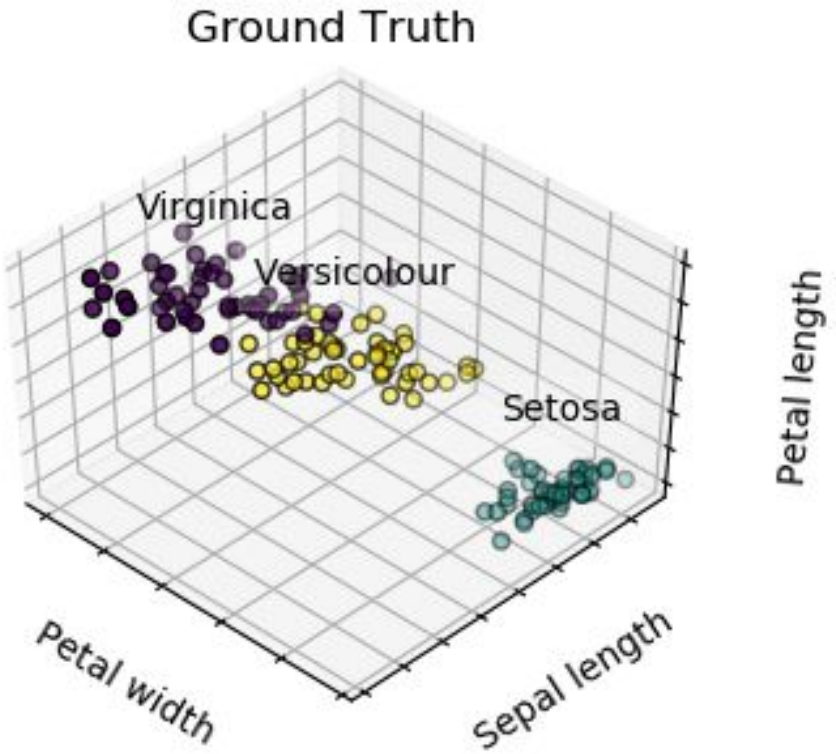
henrique.santos.003@acad.pucrs.br

joao.etchichury@acad.pucrs.br

jonatas.vargas@acad.pucrs.br


scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
predicting a category
YES
do you have labeled data
YES
<100K samples
NO
Linear SVC
NOT WORKING
Text Data
NO
Naive Bayes
YES
KNeighbors Classifier
NOT WORKING
SVC Ensemble Classifiers
NOT WORKING
SGD Classifier
NO
kernel approximation
NOT WORKING
classification
NO
number of categories known
YES
<10K samples
YES
KMeans
NOT WORKING
Spectral Clustering GMM
NO
MiniBatch KMeans
NO
<10K samples
YES
MeanShift VBGMM
NO
tough luck
clustering
NO
predicting a quantity
YES
<100K samples
NO
SGD Regressor
YES
few features should be important
YES
ElasticNet Lasso
NO
RidgeRegression SVR (kernel='linear')
NOT WORKING
SVR(kernel='rbf') EnsembleRegressors
regression
NO
just looking
YES
Randomized PCA
NOT WORKING
Isomap Spectral Embedding
NOT WORKING
LLE
YES
<10K samples
NO
kernel approximation
dimensionality reduction
NO
predicting structure

Clusterização: Iris



Classificação: Problemas x Algoritmos

A. Câncer de Mama
B. Dígitos (Imagem)
C. Vinho (3 classes)
D. Iris (3 classes)

1. K-Neighbors
2. Neural Network
3. Árvore de Decisão
4. Suport Vector Machine
5. Naive Bayes
6. Random Forest

Classificação: Problemas x Algoritmos

sklearn.datasets.

A. load_breast_cancer()
B. load_digits()
C. load_wine()
D. load_iris()

sklearn.

1. neural_network.MLPClassifier
2. neighbors.KNeighborsClassifier
3. svm.LinearSVC
4. tree.DecisionTreeClassifier
5. ensemble.RandomForestClassifier
6. naive_bayes.GaussianNB

Notebooks da Oficina

- Notebook Básico
- Métricas e Regressão Notebook
- Detecção de Outliers

Para todos os links do Colab:

- ao clicar no link, no Google Drive, ir em “Abrir com: Colaboratory”
- dentro do Google Colab, ir em “File”, depois em “Save a copy in Google Drive”
- assim será possível modificar o arquivo e rodar os experimentos

Outros Notebooks

- Regressão e Busca Exaustiva
- Processamento da Linguagem Natural
- Redes Neurais

Para todos os links do Colab:

- ao clicar no link, no Google Drive, ir em “Abrir com: Colaboratory”
- dentro do Google Colab, ir em “File”, depois em “Save a copy in Google Drive”
- assim será possível modificar o arquivo e rodar os experimentos

Outras Palestras

- Ciência de Dados na Saúde (vídeo)
- Data Science para Publicidade
- Aprendizado Não-Supervisionado e o PageRank (vídeo)

Mais material:

- https://github.com/amueller/scipy-2017-sklearn
- https://github.com/amueller/introduction_to_ml_with_python
- http://scikit-learn.org/stable/documentation.html
- https://colab.research.google.com/notebooks/basic_features_overview.ipynb
- https://www.datascienceacademy.com.br/pages/cursos-gratuitos
- https://br.udacity.com/course/intro-to-data-science--ud359

Vídeos no YouTube

- Big Data - Nerdologia
- Ciência de Dados - Nerdologia
- Machine Learning - Nerdologia
- Linguística Forense - Unabomber
- Robotização Eleitorial - Estadão
- O futuro do seu emprego - Nerdologia
- Aprendizado por Reforço - AlphaGo
- AlphaGo Zero - DeepMind
- Profissional do Futuro - TED Talks
- Why Deep Learning Now? - ColdFusion

Livros sobre Ciência de Dados (para leigos)

- Super Chunchers, Ian Ayres
- Numeratis, Stephen Baker

Datasets para Experimentos

- https://www.kaggle.com/
- https://www.openml.org/
- https://toolbox.google.com/datasetsearch
- https://mimic.physionet.org