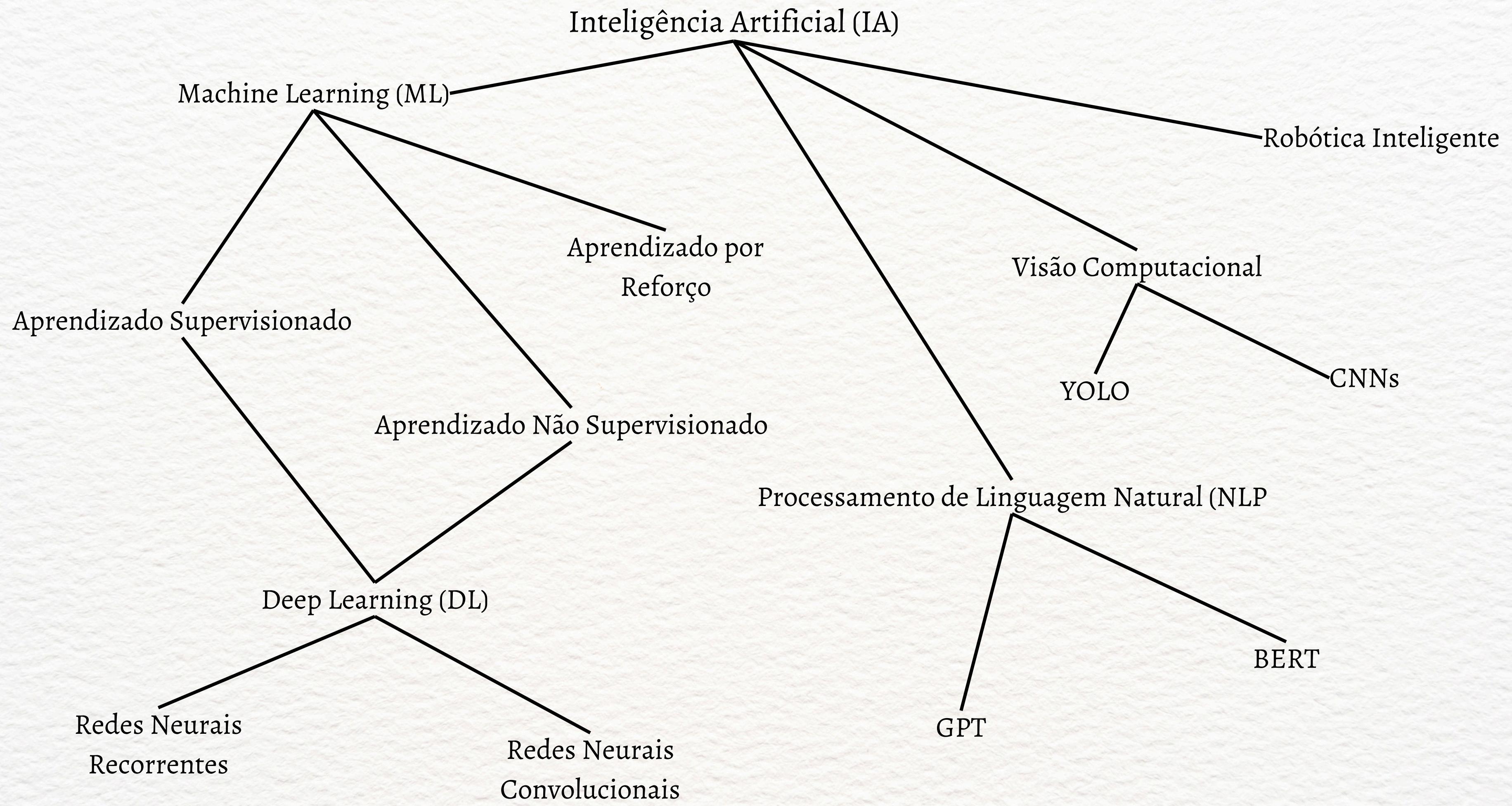


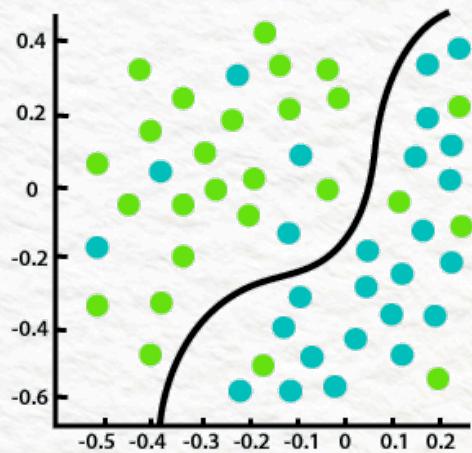
APRENDIZADO DE MÁQUINA (ML)

Henrique Poledna

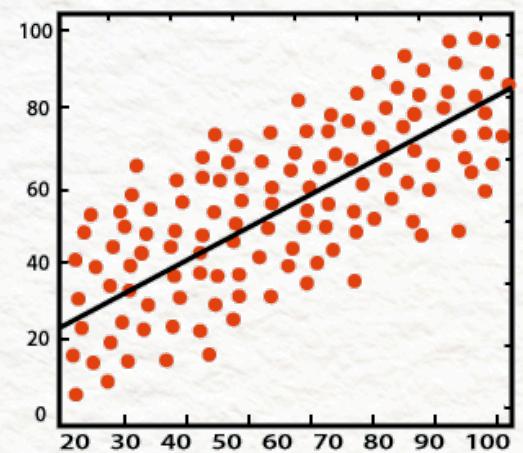
DEFINIÇÃO MACHINE LEARNING

Aprendizado de Máquina é uma das vertentes da Inteligência Artificial que permite que computadores aprendam padrões e tomem decisões a partir de dados, sem serem explicitamente programados para isso.





Classification



Regression

RNA

SUPERVISIONADO

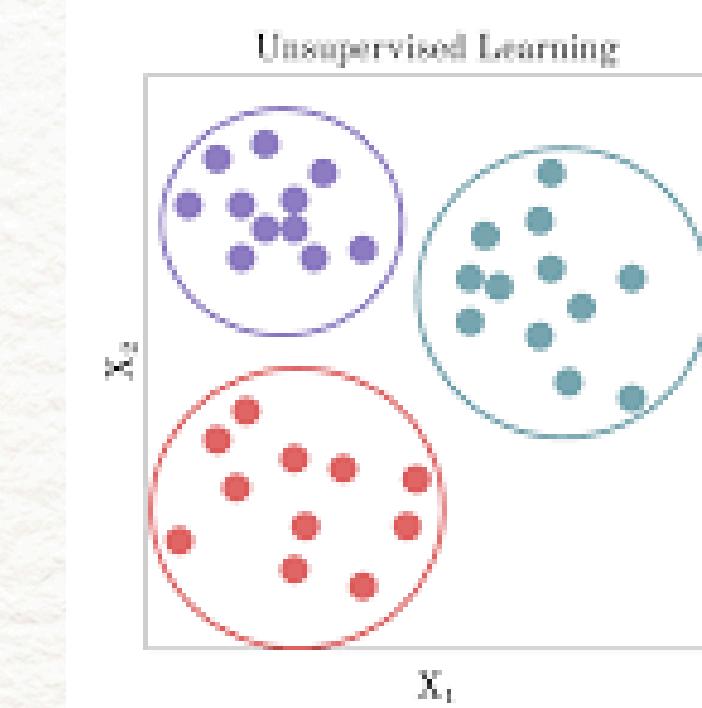
Naive Bayes

SVM

Random Forest

Árvores de Decisão

KNN



Unsupervised Learning

PCA

NÃO SUPERVISIONADO

Hierarchical Clustering

K-Means

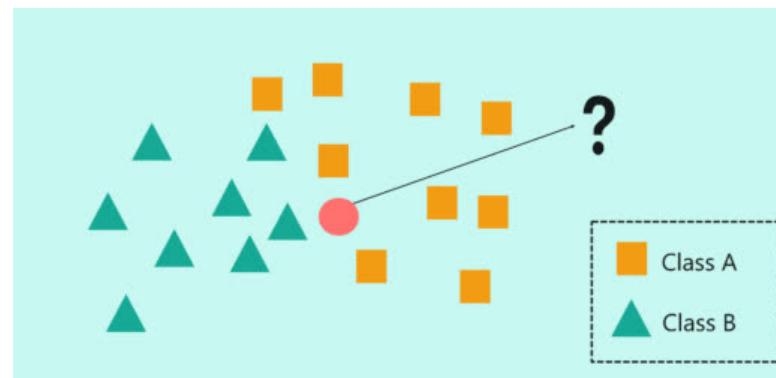
DBSCAN

DBSCAN

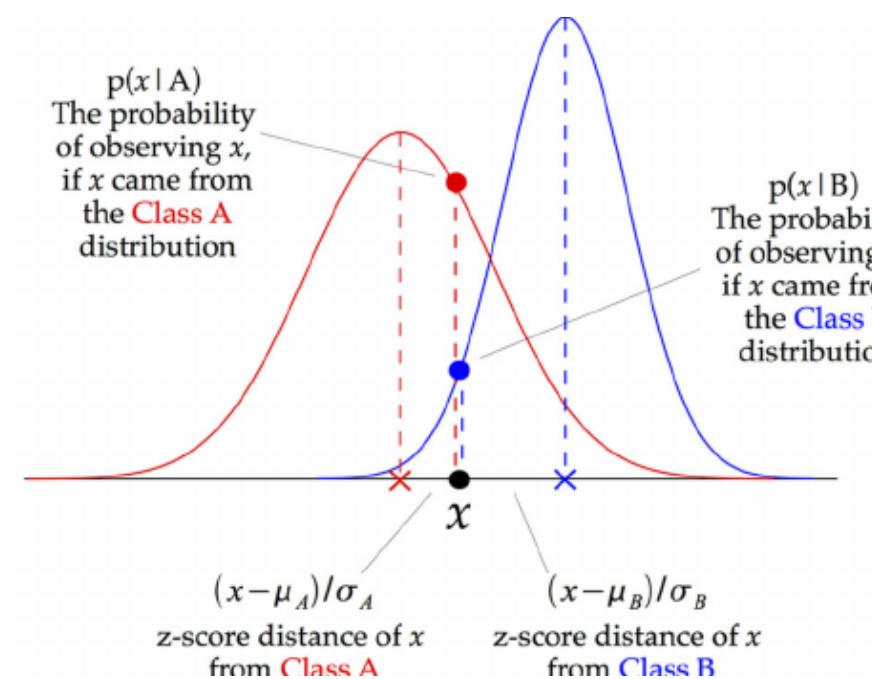
SOM

5.1,3.5,1.4,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
5.7,2.8,4.5,1.3,Iris-versicolor
7.6,3.0,6.6,2.1,Iris-virginica
4.9,2.5,4.5,1.7,Iris-virginica
6.4,2.7,5.3,1.9,Iris-virginica

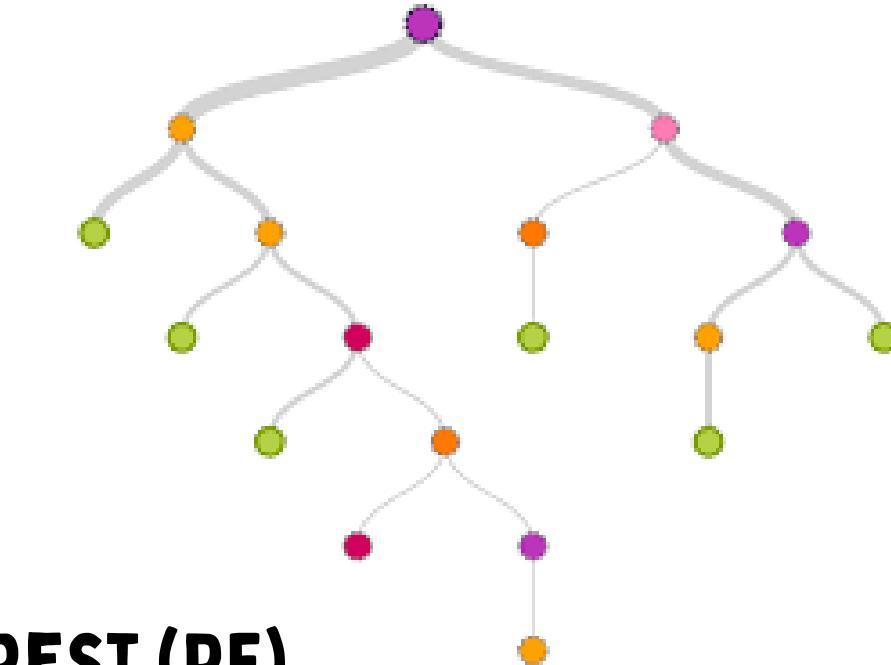
KNN



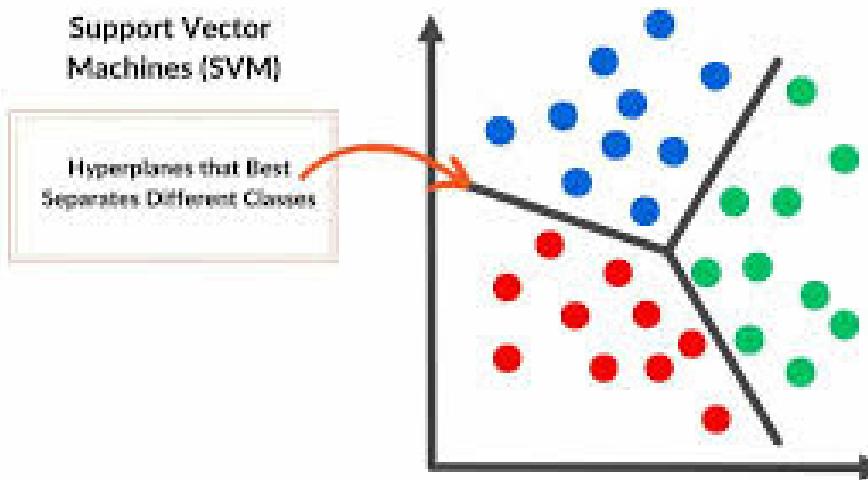
NAIVE BAYES



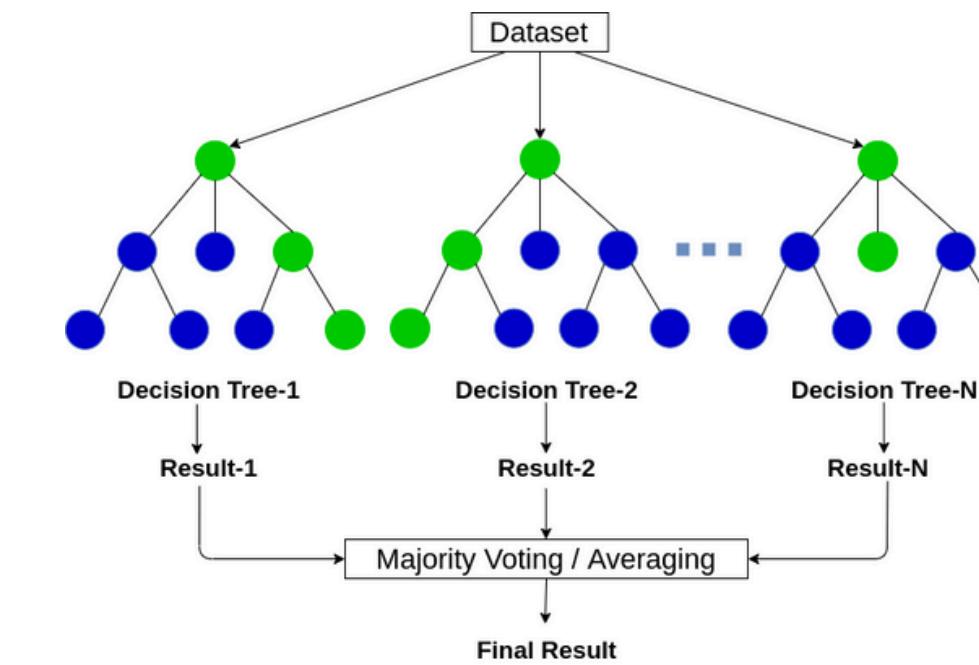
ÁRVORE DE DECISÃO



SUPPORT VECTOR MACHINE (SVM)

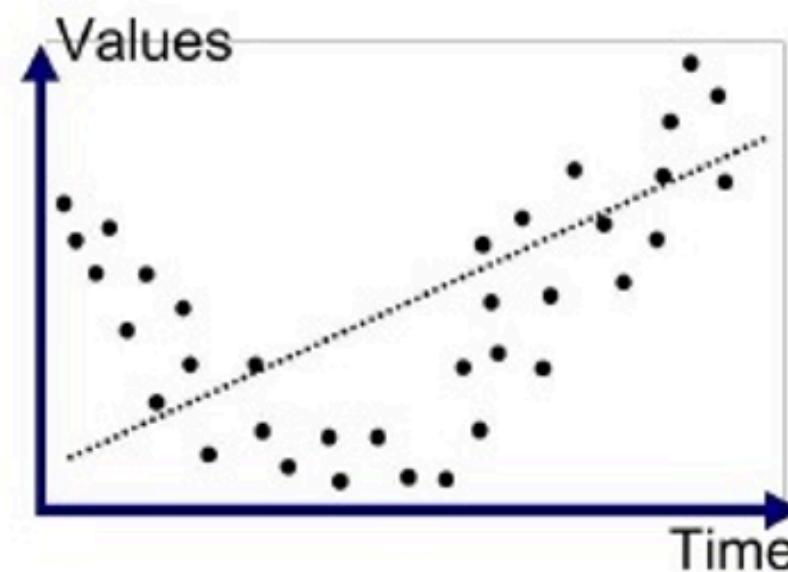


RANDOM FOREST (RF)

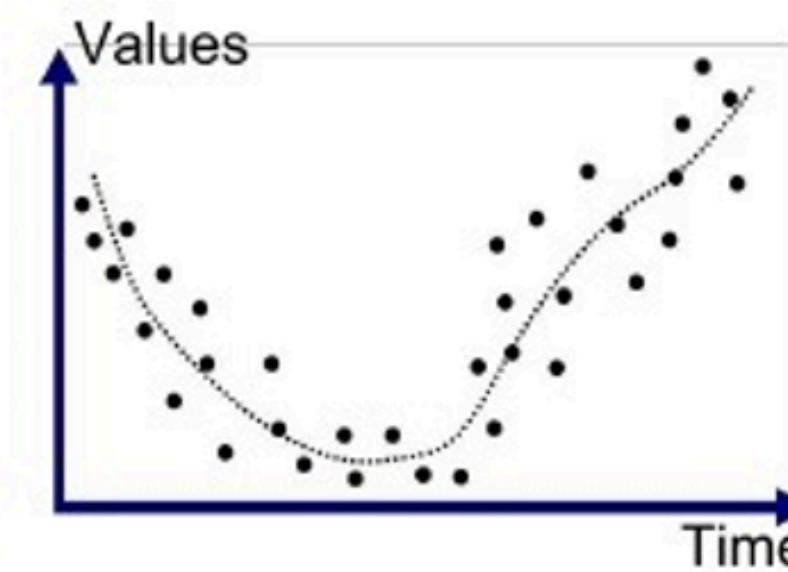


AVALIAÇÃO DE MODELOS

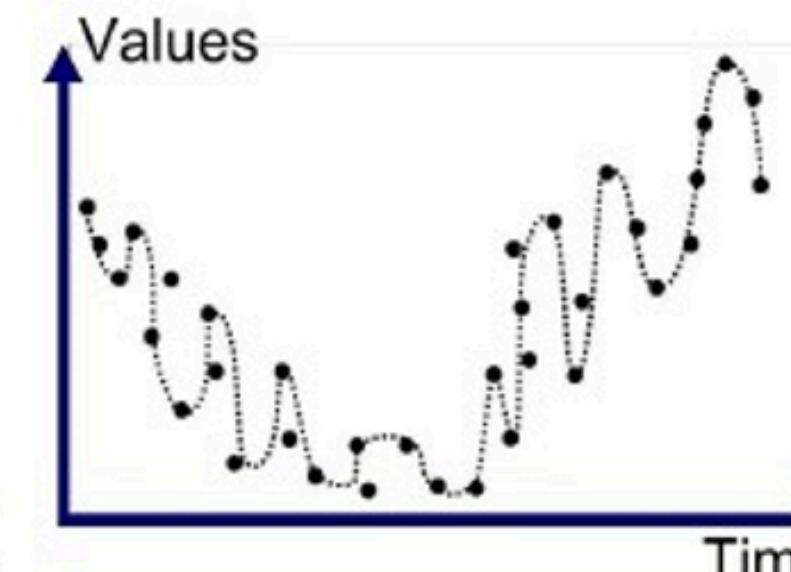
UNDERFITTING E OVERFITTING



Underfitted



Good Fit/R robust



Overfitted

EXEMPLO

Pressão Sanguínea (mmHg)	Concentração de Glicose (mg/dl)	Idade (anos)	IMC (Kg/m ²)	Target
72	148	47	43.6	Sim
63	87	33	28.4	Não
67	182	36	25.2	Sim
64	90	22	29.1	Não
...

MATRIZ DE CONFUSÃO

		REAL	
		POSITIVO	NEGATIVO
--- PREVISTO ---	POSITIVO	VP	FP
	NEGATIVO	FN	VN

**AO UTILIZAR O MODELO SVM,
OBTIVEMOS ESTA MATRIZ.**

-- PREVISTO --

----- REAL -----

	COM DIABETES	SEM DIABETES
COM DIABETES	243	34
SEM DIABETES	25	465

E AO UTILIZAR O MODELO RF, OBTIVEMOS
ESTA MATRIZ.

		----- REAL -----	
		COM DIABETES	SEM DIABETES
-- PREVISTO --	COM DIABETES	213	64
	SEM DIABETES	55	435

**O SVM ACERTOU MAIS AO
CLASSIFICAR PACIENTES COMO
DIABÉTICOS (VP), COM 243
ACERTOS, CONTRA 213 DO RF.
CONSEQUENTEMENTE TAMBÉM
ERROU MENOS AO TER 34
CLASSIFICAÇÕES (CONTRA 64 DO
RF) COMO DIABÉTICOS, QUE NA
VERDADE NÃO ERAVAM DIABÉTICOS
(FP)**

**O SVM ACERTOU MAIS AO
CLASSIFICAR PACIENTES COMO
NÃO DIABÉTICOS (VN), COM 465
ACERTOS, CONTRA 435 DO RF.
CONSEQUENTEMENTE TAMBÉM
ERROU MENOS AO TER 25
CLASSIFICAÇÕES (CONTRA 55 DO
RF) COMO NÃO DIABÉTICOS, QUE
NA VERDADE ERAVAM DIABÉTICOS
(FN)**

ACCURACY OU ACURÁCIA

A ACURÁCIA É UMA MÉTRICA MUITA BOA QUANDO A BASE DE DADOS ESTÁ BALANCEADA

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}$$

SVM - VP = 243; VN = 465; FN = 25 E FP = 34 R = 0,92

RF - VP = 213; VN = 435; FN = 55 E FP = 64 R = 0,84

DE TODAS AS CLASSIFICAÇÕES QUE O MODELO REALIZOU, O QUANTO ELE ACERTOU?

PRECISION OU PRECISÃO

$$\text{Precisão} = \frac{VP}{VP + FP}$$

SVM - VP = 243 E FP = 34 R = 0,88

RF - VP = 213 E FP = 64 R = 0,77

DE TODAS AS PREVISÕES QUE O MODELO CLASSIFICOU COMO POSITIVO, QUANTAS ELE ACERTOU?

RECALL OU REVOCAÇÃO

Resumindo, a revocação mede a capacidade do modelo de evitar os FN

$$\text{Revocação} = \frac{VP}{VP + FN}$$

SVM - VP = 243 E FN = 25 R = 0,91

RF - VP = 213 E FP = 55 R = 0,79

DE TODOS OS CASOS QUE SÃO REALMENTE POSITIVOS, O QUANTO O MODELO ACERTOU?

F1-SCORE OU MEDIDA-F1

$$F1 = 2 \times \frac{Precisão \times Revocação}{Precisão + Revocação}$$

SVM - P = 20,88 E R = 0,91 R = 0,88

RF - P = 0,77 E R = 0,79 R = 0,78

A MEDIDA-F1 (F1-SCORE) NADA MAIS É QUE UMA MÉDIA HARMÔNICA ENTRE A PRECISÃO E A REVOCAÇÃO

REFERÊNCIAS

- <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier>
- <https://pt.khanacademy.org/math/pre-algebra/pre-algebra-equations-expressions/pre-algebra-dependent-independent/a/dependent-and-independent-variables-review>
- <https://didatica.tech/scikit-learn-na-pratica-codigos-uteis-e-comandos-essenciais/>
- <https://didatica.tech/dados-de-treino-e-teste/>
- <https://sigmoidal.ai/metricas-de-avaliacao-em-modelos-de-classificacao-em-machine-learning/>
- <https://learn.microsoft.com/pt-br/shows/machine-learning-for-beginners/introduction-to-machine-learning-for-beginners-machine-learning-for-beginners>
- <https://www.youtube.com/watch?v=IHAb3NHDahU>
- <https://developers.google.com/machine-learning/intro-to-ml/what-is-ml?hl=pt-br>
- <https://www.youtube.com/watch?v=N2TT2Q83abc>
- https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation
- <https://sigmoidal.ai/metricas-de-avaliacao-em-modelos-de-classificacao-em-machine-learning/>
- <https://cloud.google.com/learn/what-is-machine-learning?hl=pt-BR>
- <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://optimus-solutions.com.br/analise-de-dados-machine-learning/&ved=2ahUKEwjaqMnlka2MAxUqkZUCHV39HBMQFnoECDAQAQ&usg=AOvVaw3nicAc7iefUrIY4CeNCFbC>
- <https://scikit-learn.org/stable/modules/neighbors.html#classification>
- <https://medium.com/@nirajan.acharya777/understanding-precision-recall-f1-score-and-support-in-machine-learning-evaluation-7ec935e8512e>
- <https://gist.github.com/tijptjik/9408623>
- <https://github.com/GustavoAkyama/wine-quality-predict>
- <https://www.youtube.com/watch?v=FZqMCgCb03U>
- <https://blog.mettzer.com/variaveis-dependentes-e-independentes/>
- <https://www.ibm.com/br-pt/topics/machine-learning>
- https://www.sas.com/pt_br/insights/analytics/machine-learning.html
- <https://medium.com/data-hackers/principais-m%C3%A9tricas-de-classifica%C3%A7%C3%A3o-de-modelos-em-machine-learning-94eeb4b40ea9>
- <https://www.datacamp.com/tutorial/precision-recall-curve-tutorial>

