

# DATASETs

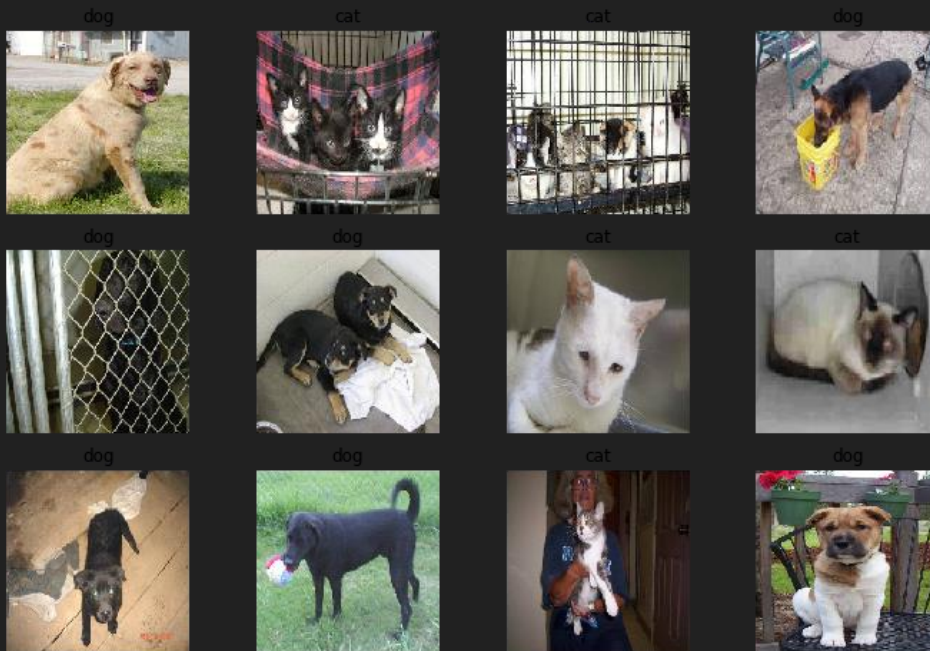
## Teoria e Prática

**Prof. Dr. Diego Bruno**

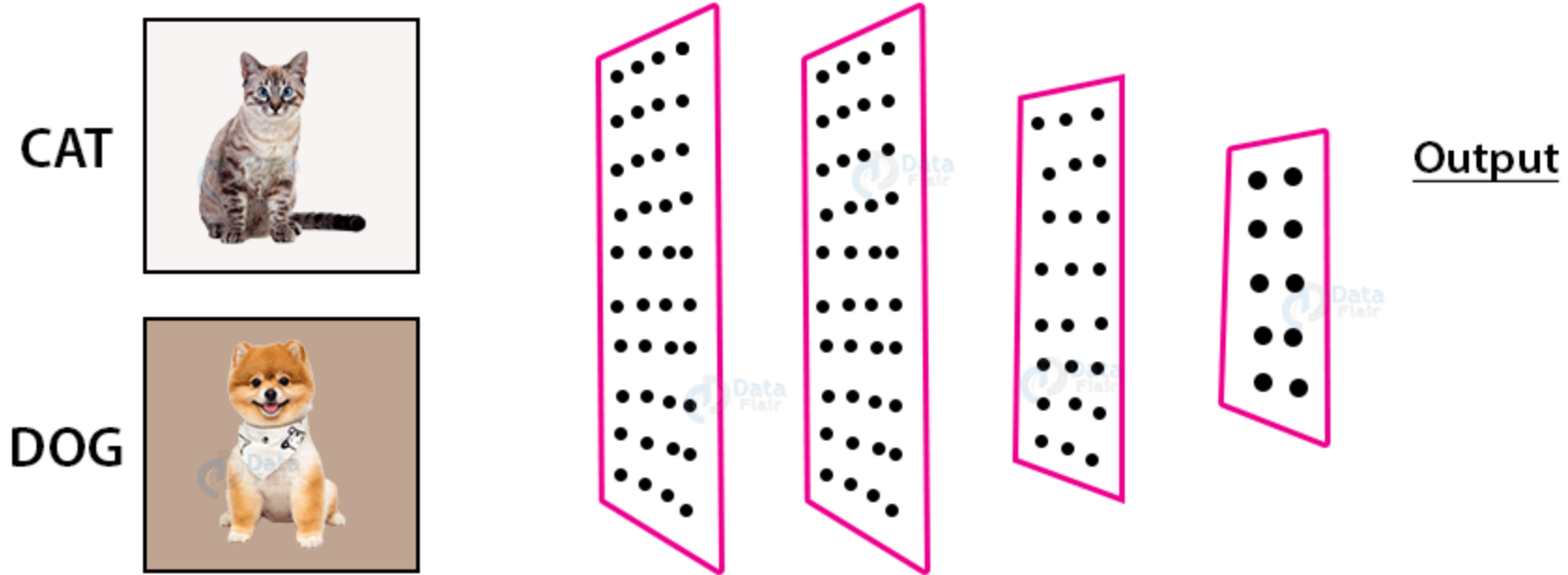
Education Tech Lead na DIO

Doutor em Robótica e *Machine Learning* pelo ICMC-USP

# O que é um Dataset?



# Dataset



# Dataset

## Como devem ser minhas amostras?

Predicted:Dog



Predicted:Cat



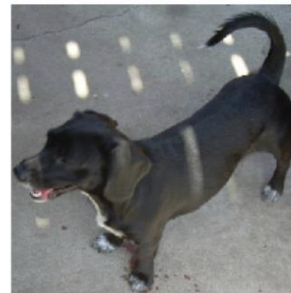
Predicted:Cat



Predicted:Dog



Predicted:Dog



Fonte: [https://www.linkedin.com/pulse/cats-vs-dogs-image-classification-using-cnn-piyush-pareek/?trk=public\\_profile\\_article\\_view](https://www.linkedin.com/pulse/cats-vs-dogs-image-classification-using-cnn-piyush-pareek/?trk=public_profile_article_view)

# Dataset

Como devem ser minhas amostras?

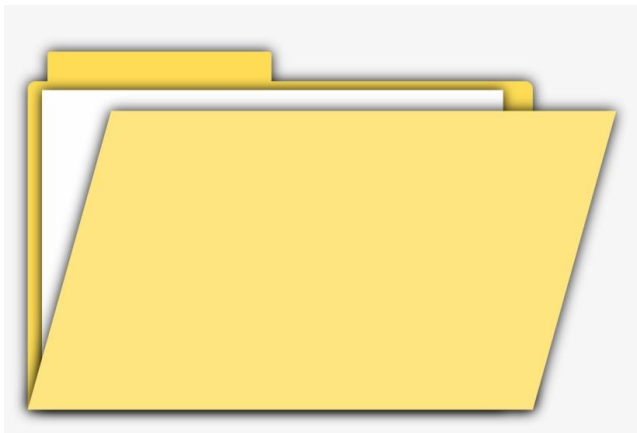


Fonte: [https://www.linkedin.com/pulse/cats-vs-dogs-image-classification-using-cnn-piyush-pareek/?trk=public\\_profile\\_article\\_view](https://www.linkedin.com/pulse/cats-vs-dogs-image-classification-using-cnn-piyush-pareek/?trk=public_profile_article_view)

# Como criar um Dataset?



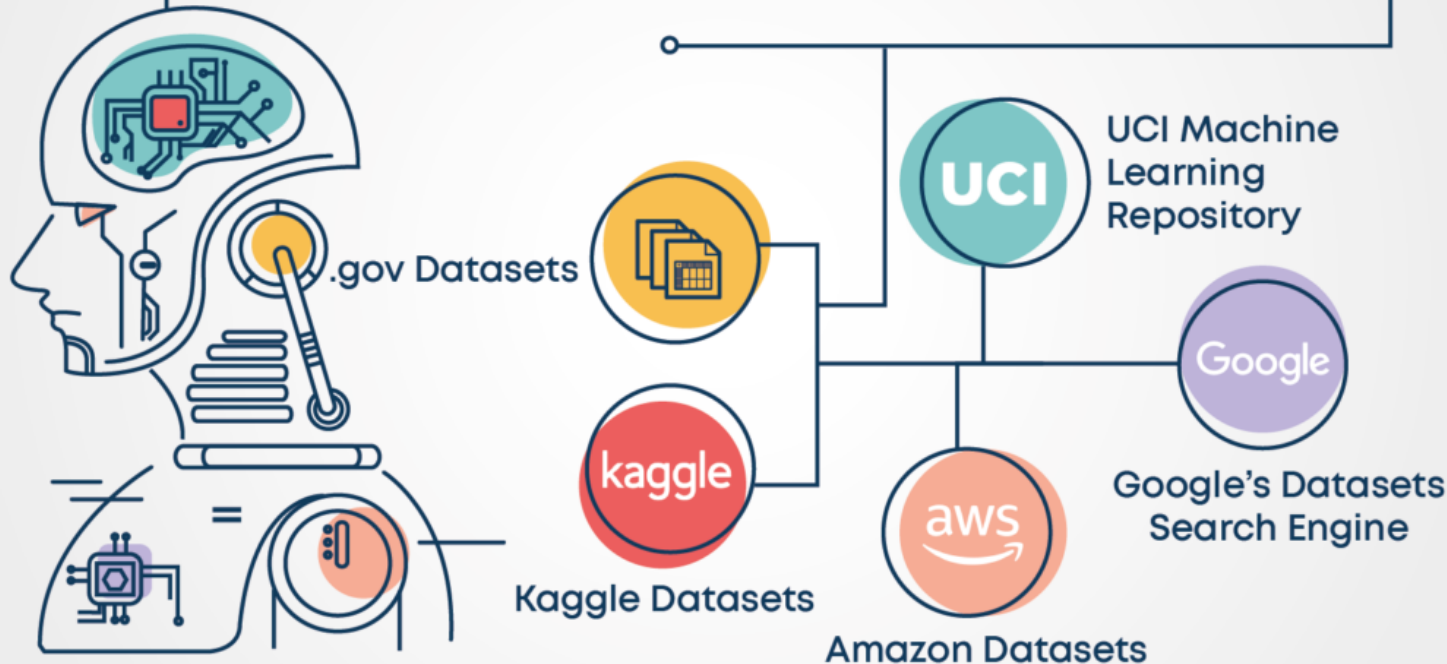
Como gerar uma base de dados...



# Serviços de DATASETs?

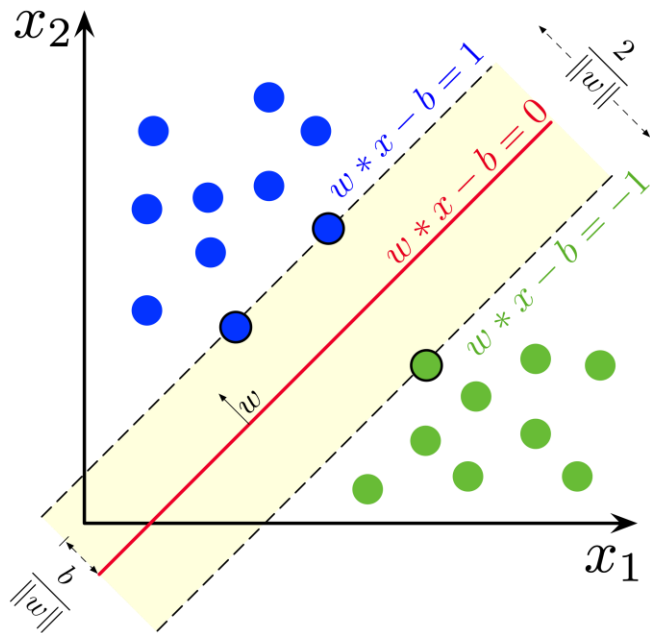
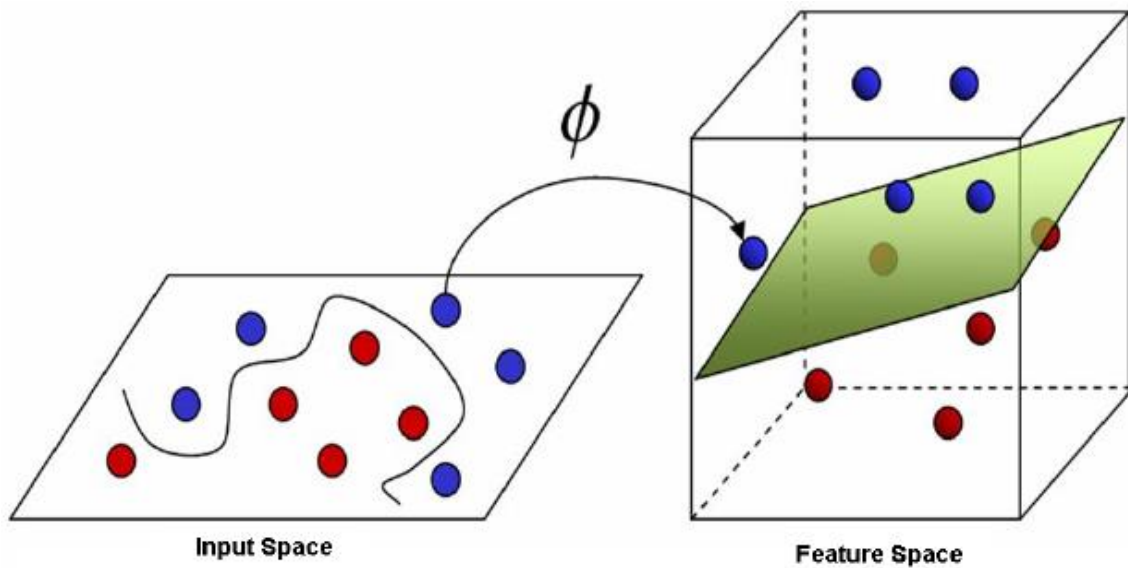
## Bases de dados disponíveis

### TOP 5 SOURCES FOR MACHINE LEARNING AND ANALYTICS DATASETS



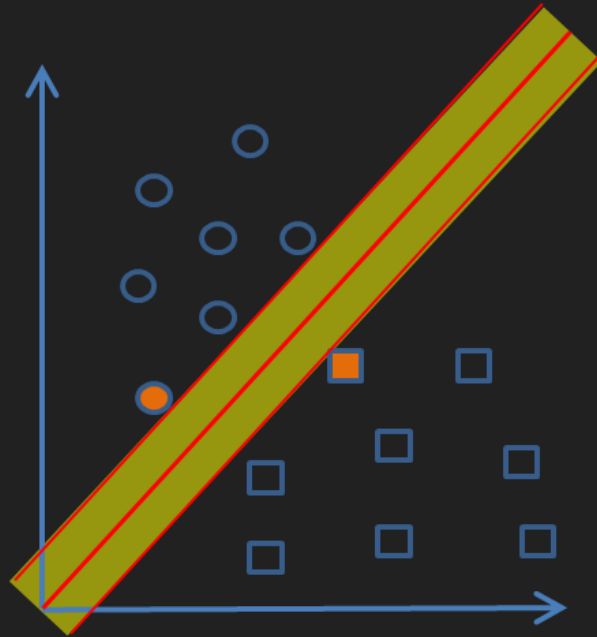
# O que são SVMs?

## Máquina de Vetores de Suporte



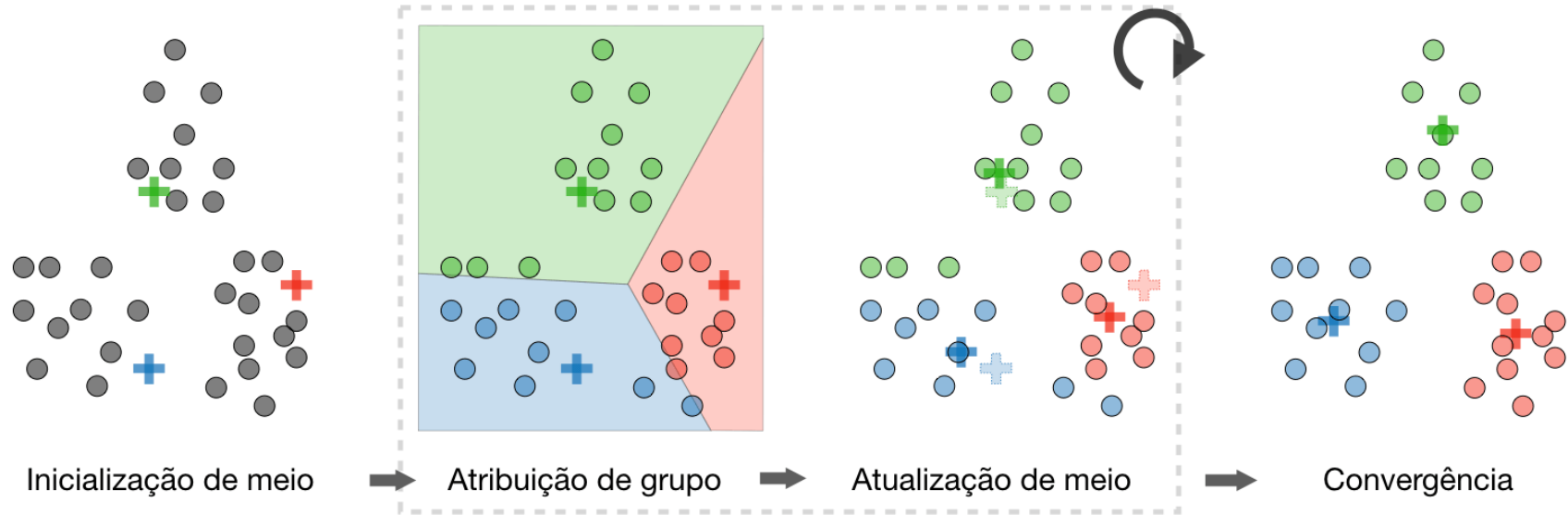


# Tipos de aprendizado



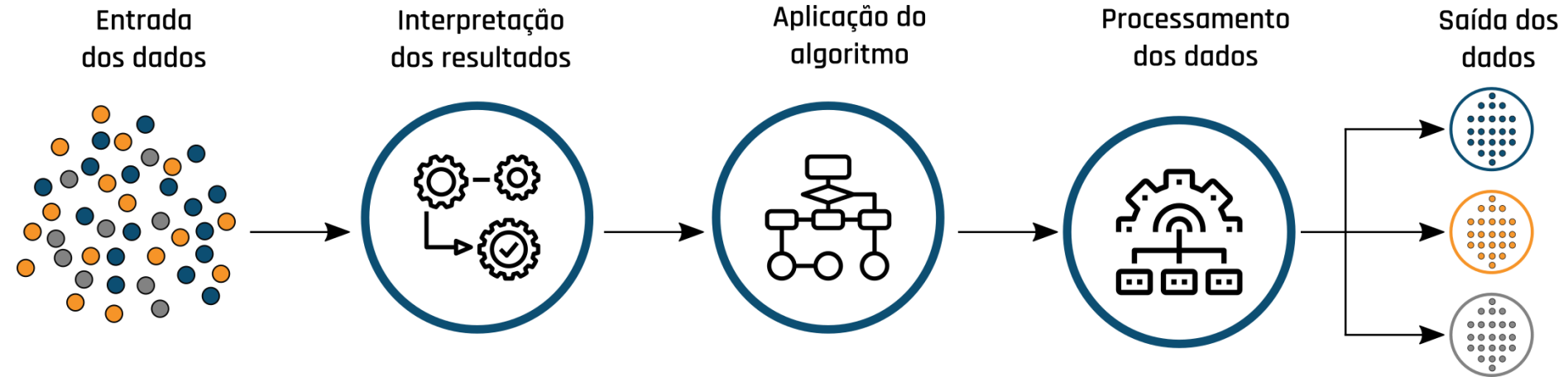
# Tipos de aprendizado

## Não supervisionado



# Tipos de aprendizado

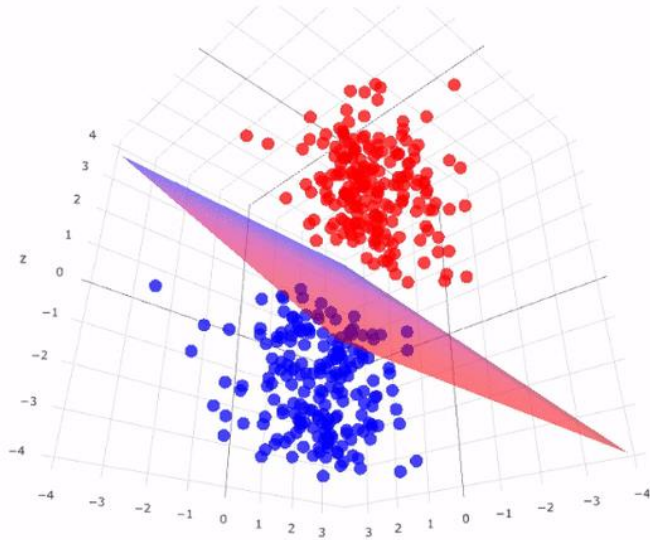
## Supervisionado



# Tipos de aprendizado

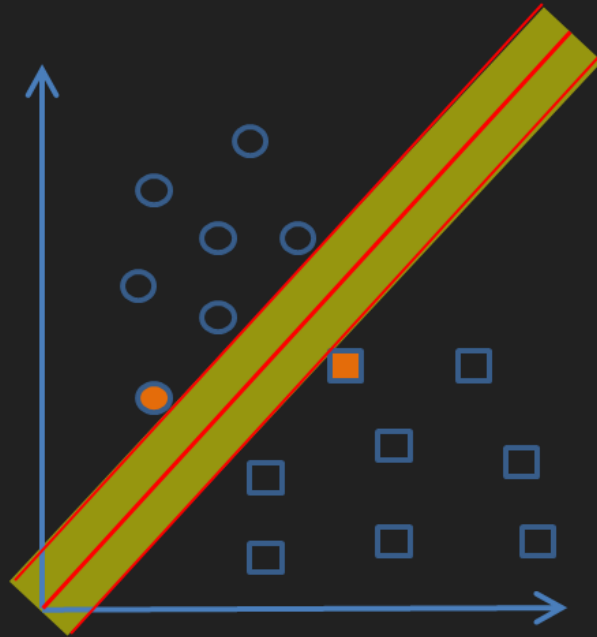


## Supervisionado



Os **algoritmos** de aprendizagem **supervisionada** relacionam uma saída com uma entrada com base em dados rotulados. Neste caso, o usuário alimenta ao **algoritmo** pares de entradas e saídas conhecidos.

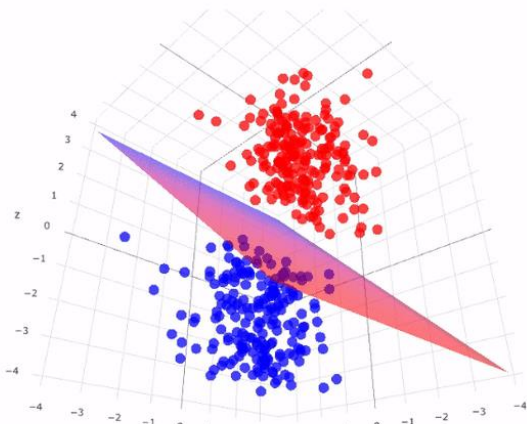
# Diferença entre RNA e SVM



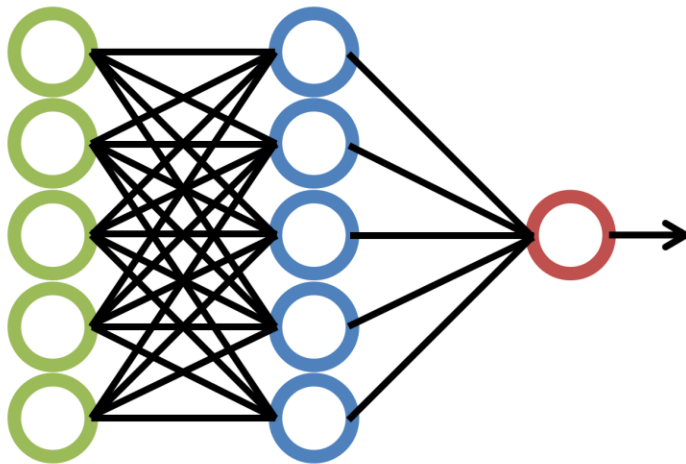
# Diferenças entre RNA e SVM?



## Supervisionado



**SVM**



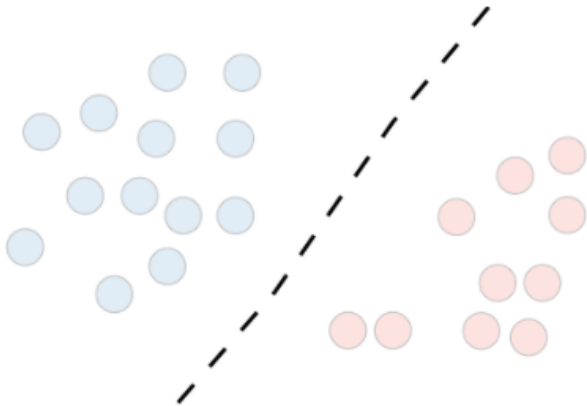
**RNA**

Na prática não há muita diferença... O principal fator é o modo de estabelecer o **hiperplano**.

**SVM** buscando a otimização das margens e a **RNA** buscando o mínimo global

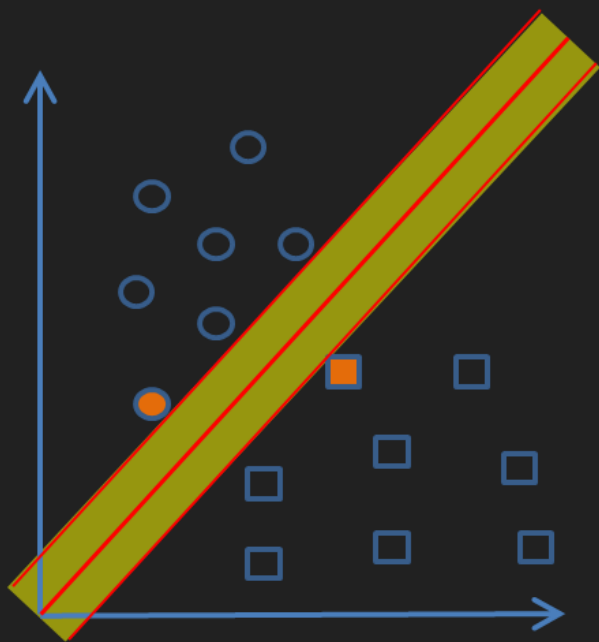
# Resultado esperado de uma SVM



	Modelo discriminativo
Objetivo	Estimar diretamente $P(y x)$
O que é aprendido	Fronteira de decisão
Ilustração	
Exemplos	Regressões, SVMs

**SVM** buscando a otimização das margens e a **RNA** buscando o mínimo global

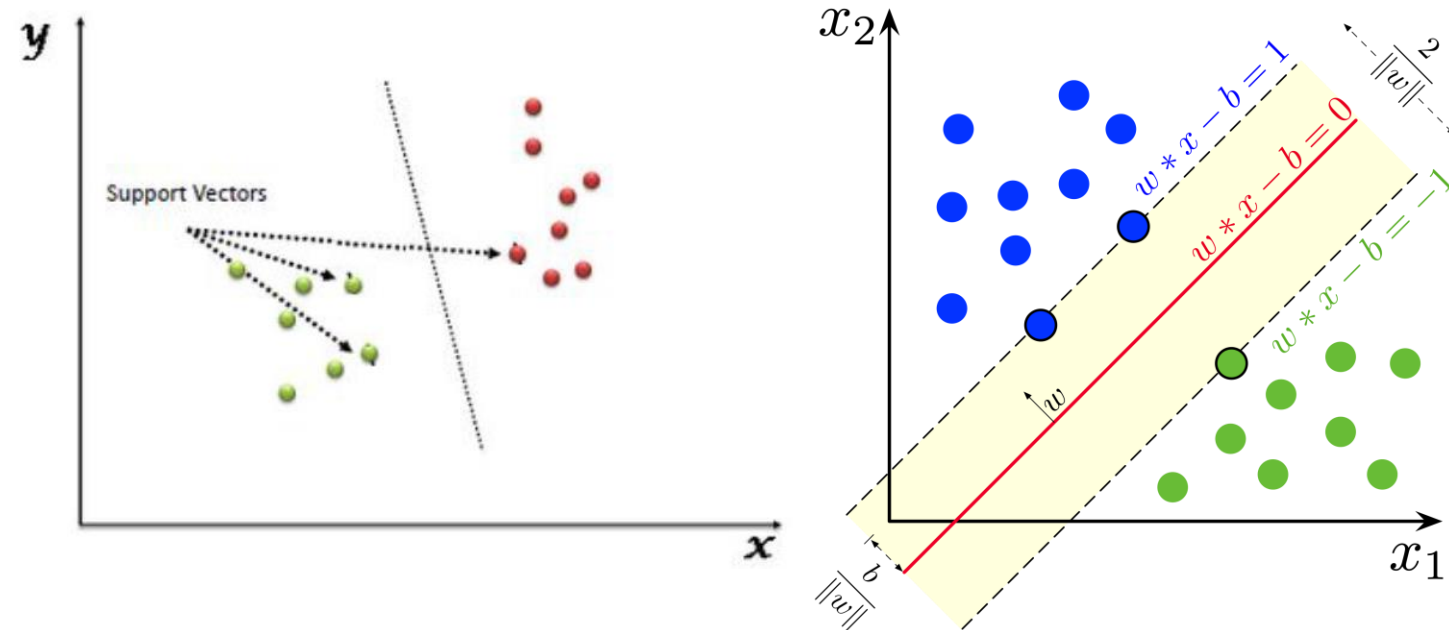
# Por que “Vetores”?



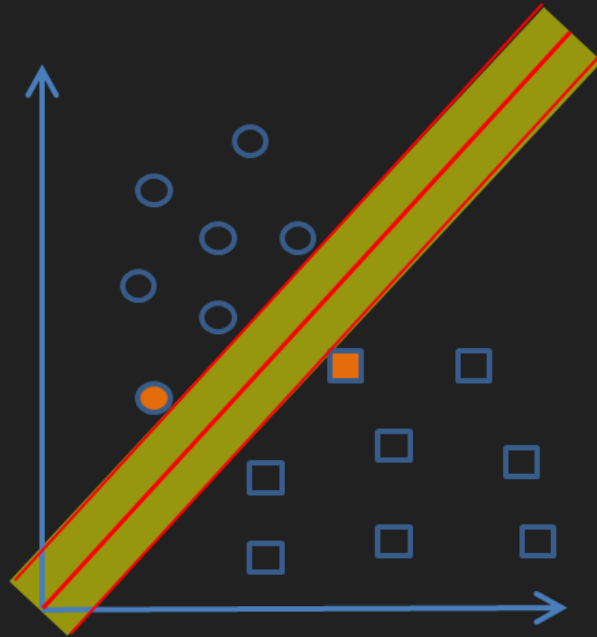


# Por que “Máquina de Vetores”?

Os “**Vetores de suporte**” são simplesmente as coordenadas da observação individual. Uma **SVM** é uma fronteira que melhor realiza as duas classes (hiperplano / linha).

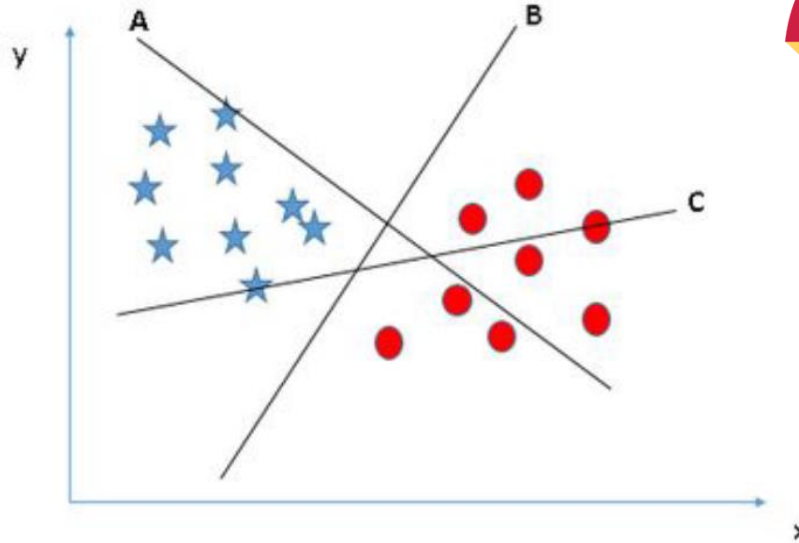


# Desenvolvimento



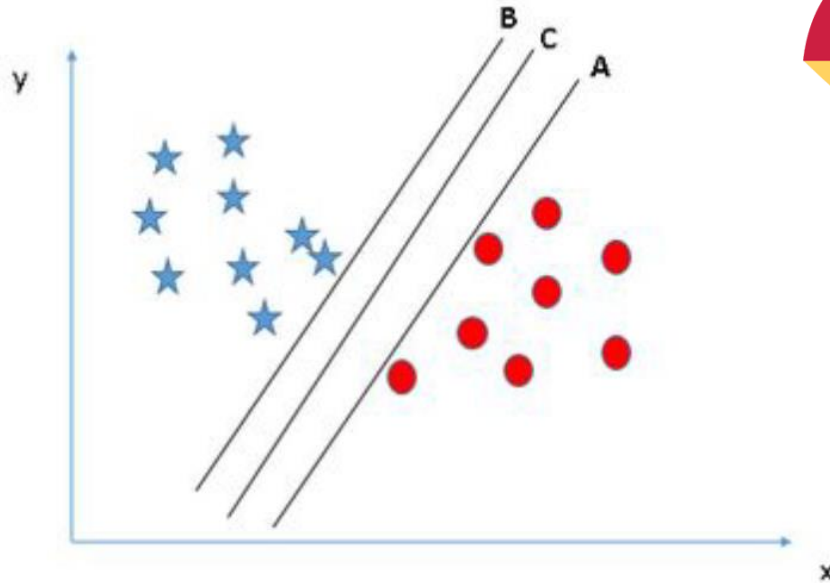
# Desenvolvendo a hipótese

**Desenvolvendo a hipótese:** Aqui, temos três hiperplanos (A, B e C). Mas qual o hiperplano certo para classificar estrela e círculo?



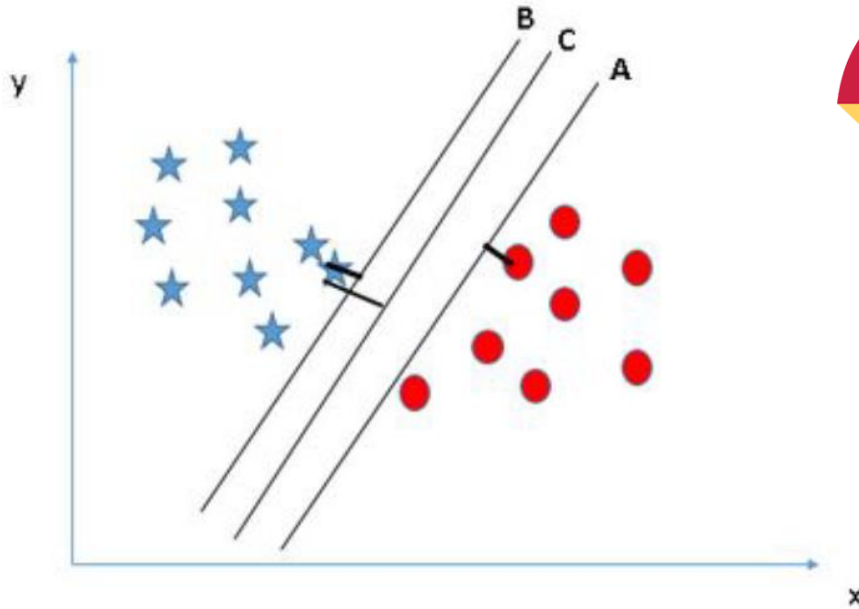
# Desenvolvendo a hipótese

Aqui, temos três **hiperplanos (A, B e C)** e todos estão dividindo bem as classes. Agora, como podemos identificar o hiperplano certo?



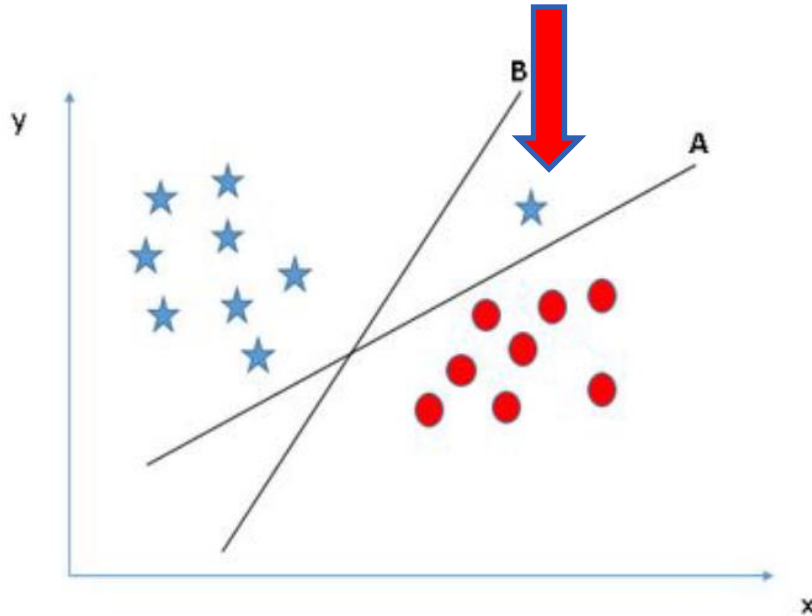
# Desenvolvendo a hipótese

**Maximizar** as distâncias entre o ponto de dados mais próximo (de qualquer classe) e o hiperplano nos ajudará a decidir o hiperplano correto.



# Desenvolvendo a hipótese

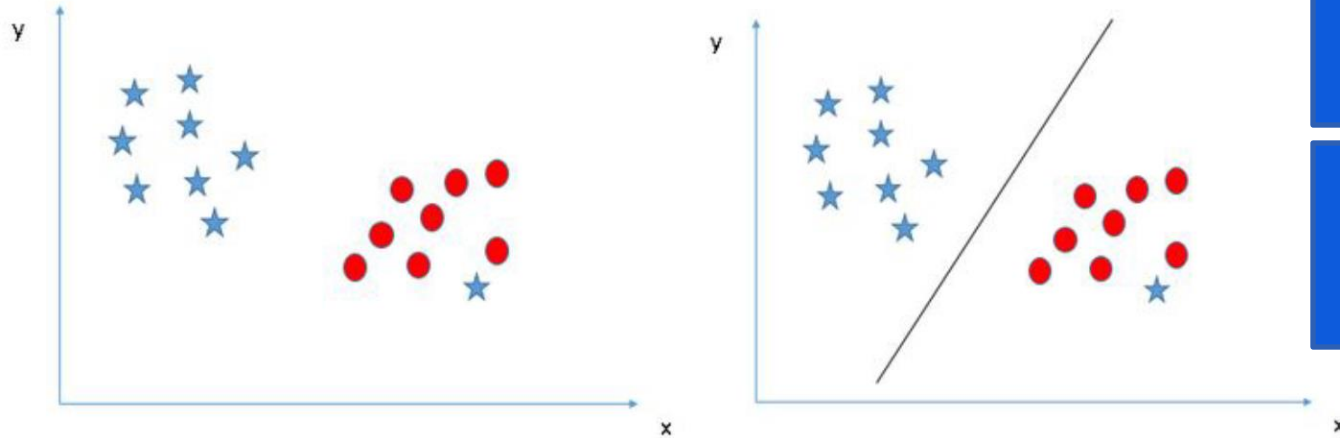
Neste caso, o melhor hiperplano é o B (já que ele tem uma margem maior em comparação a A)?



# Desenvolvendo a hipótese



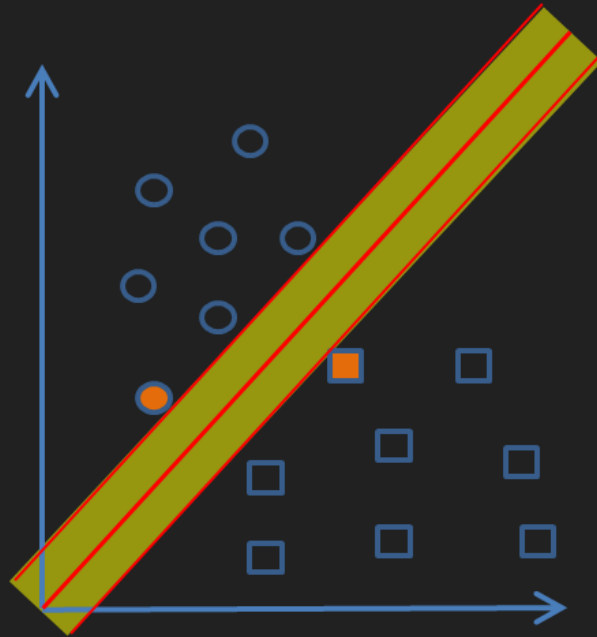
Existem caso onde não é possível separar as duas classes usando uma linha reta, pois uma das classes está no território de outra (**outlier**).



O **SVM** tem recursos para ignorar valores discrepantes

**SVM** é robusto para *outliers*

# Algoritmos





# Obrigado!

Prof. Dr. Diego Bruno