09\_07\_2020

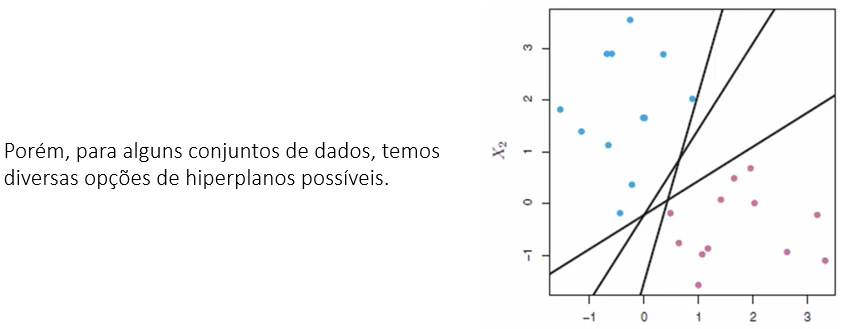
SVM

**Dados Lineares**

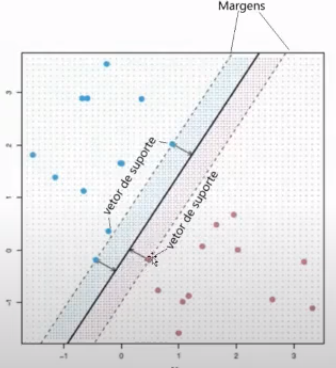
Support vector machines SVMs são métodos de machine learning baseados em aprendizado supervisionado que analizam dados e reconhecem padrões, usados para classificação e regressão.

Dado um conjunto de dados de treino, cada um marcado com uma classe pertencendo a uma categoria SVM cria um modelo que assimila novos exemplos em uma das categorias, fazendo uso de um classificador não probabilístico.

Bola azul tumor benigno – bola vermelha tumor maligno

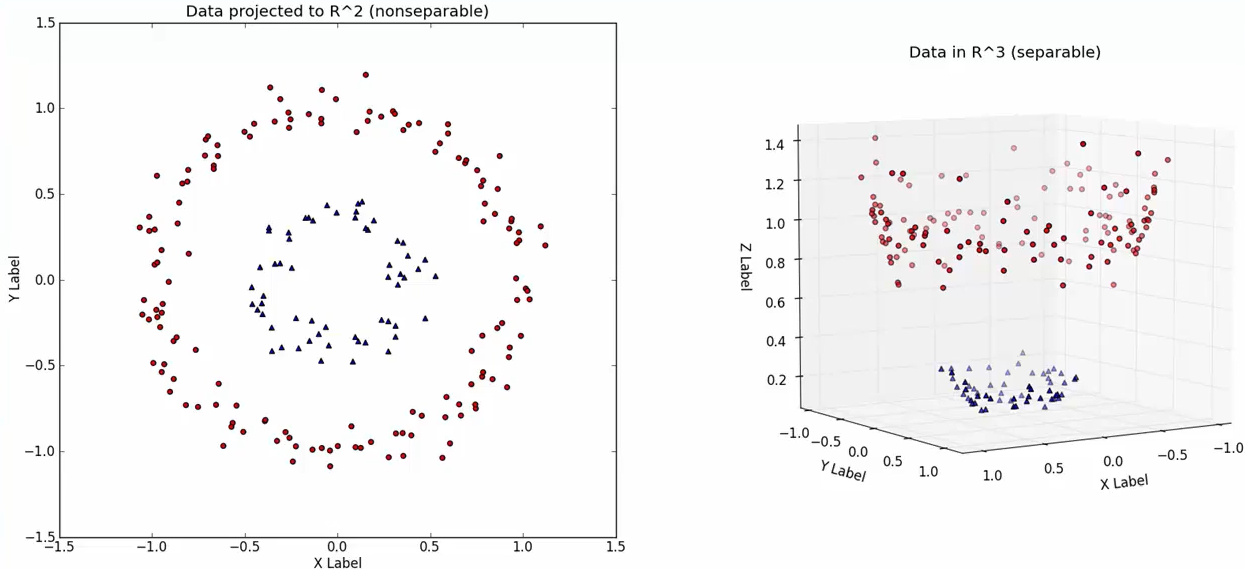


Os pontos que tocam as linhas das margens são conhecidos como vetores de suportes (Support Vectors)

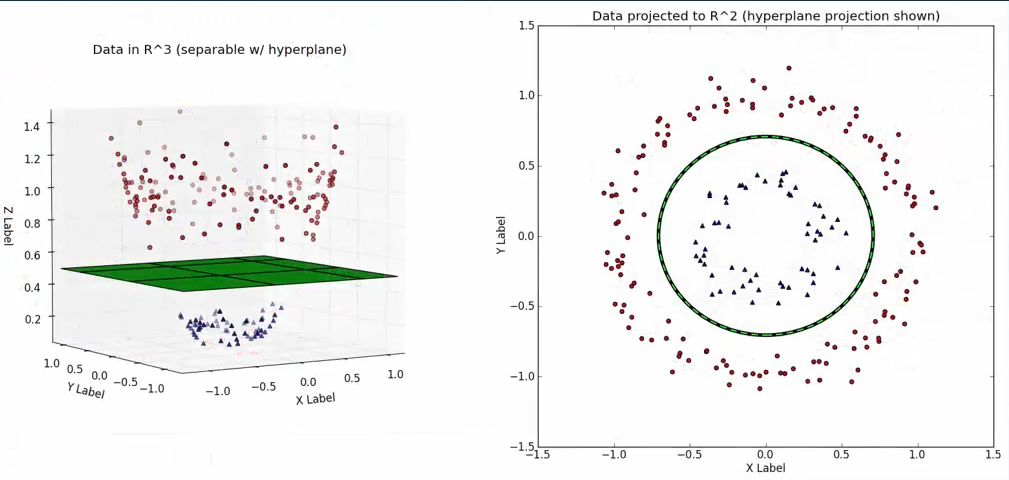


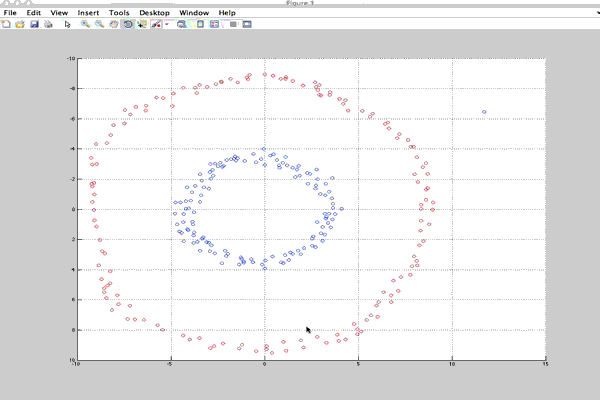
Podemos expandir esse conceito para dados não linearmente separáveis através do “**kernel trick**”

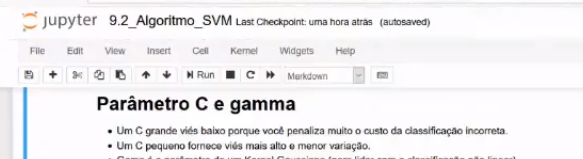
**Dados Não Lineares**



O **kernel trick** se baseia em aplicar uma transformação nos dados e leva-los para um espaço dimensional superior onde lá ele possa ser separado linearmente e depois, os trazemos de volta aplicando a inversa.







SVM

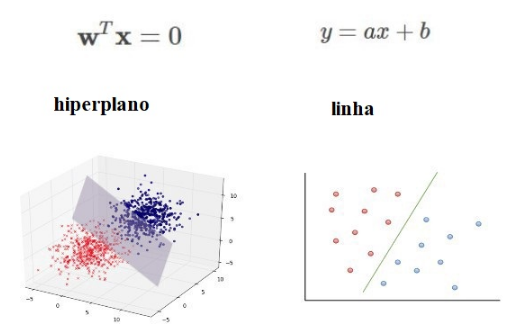
SVM é um algoritmo construído para **encontrar a melhor fronteira de decisão** que separa duas únicas classes (A e B) a **partir das amostras de treinamento**.

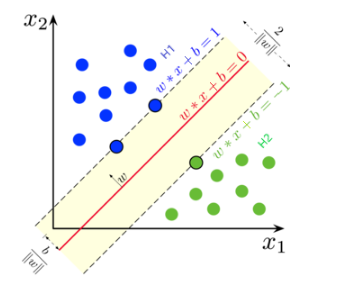
Ele faz isso por um processo de otimização contínua em que busca encontrar o hiperplano que mais bem separa as duas classes. Esse hiperplano é aquele cuja distância dos exemplos das classes é a maior possível (de modo a ficar o mais bem separado possível!). Nem todos os exemplos das classes A e B são necessários para se calcular essa distância e obter esse hiperplano, e são justamente esses pontos (ou vetores, em termos geométricos utilizados na fundamentação matemática do algoritmo) que importam. Eles são chamados de "vetores de suporte" porque caracterizam o hiperplano separador, e é daí justamente que vem o nome do algoritmo.

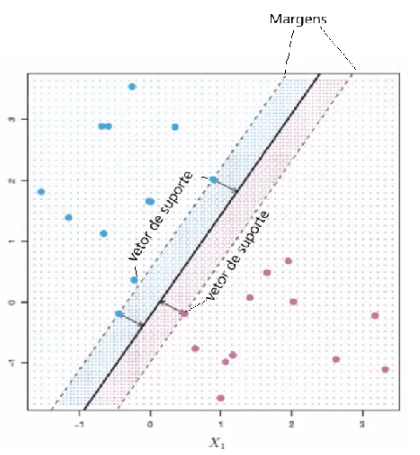
Em outras palavras, o que uma SVM faz é encontrar uma linha de separação, mais comumente chamada de hiperplano entre dados de duas classes. Essa linha busca maximizar a distância entre os pontos mais próximos em relação a cada uma das classes.

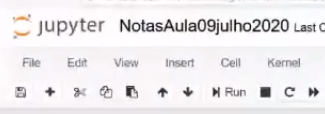
Essa distância entre o hiperplano e o primeiro ponto de cada classe costuma ser chamada de margem. A SVM coloca em primeiro lugar a classificação das classes, definindo assim cada ponto pertencente a cada uma das classes, e em seguida maximiza a margem. Ou seja, ela primeiro classifica as classes corretamente e depois em função dessa restrição define a distância entre as margens.

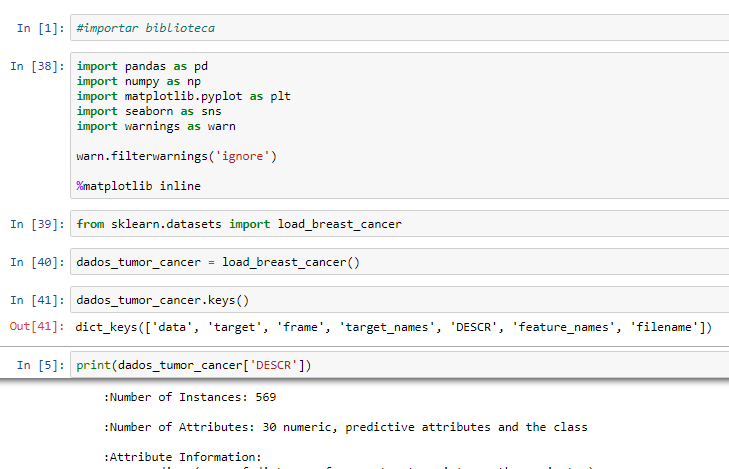
Algoritmo



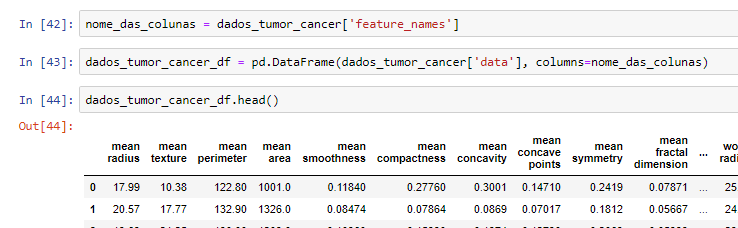


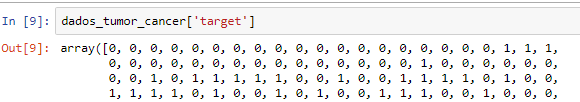


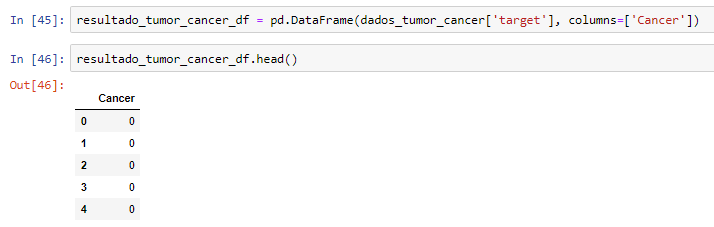




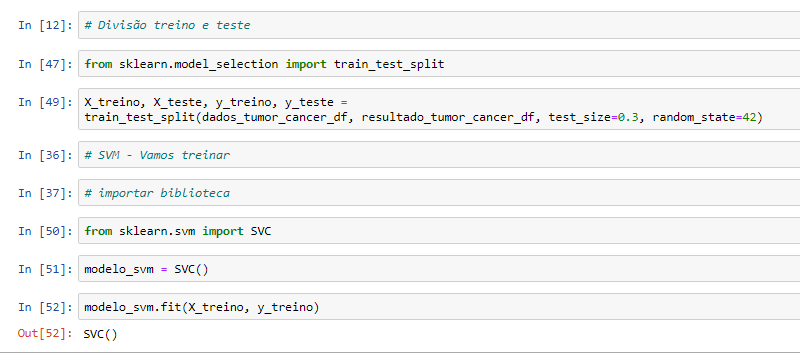
Pequena exploração dos dados



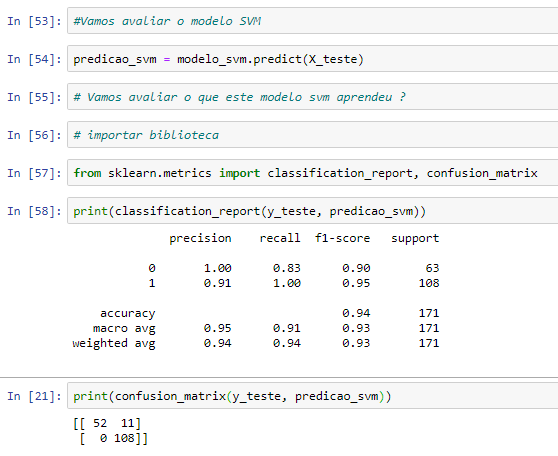




Divisão de treino e teste

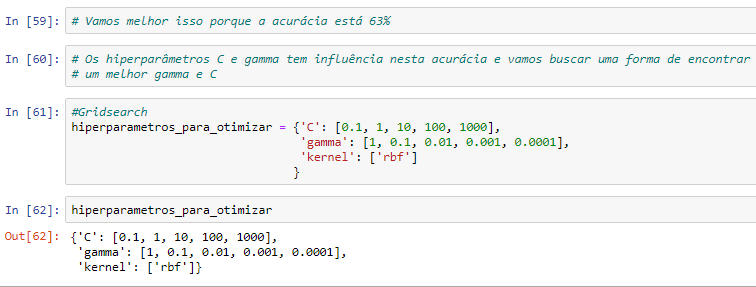


Avaliação treino, teste e Matriz de confusão

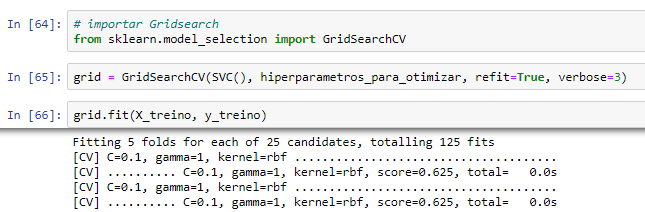


Melhorar com GridSearchCV

Hiperparametros C, Gamma e Kernel



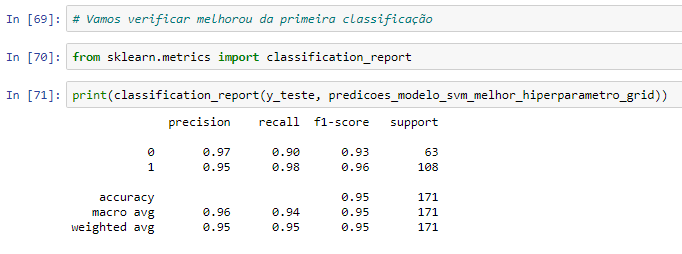
Importar Gridsearch





Predições





Novo Modelo

