



# **SUPPORT VECTOR MACHINES (SVM)**

---

# SUPPORT VECTOR MACHINES (SVM)

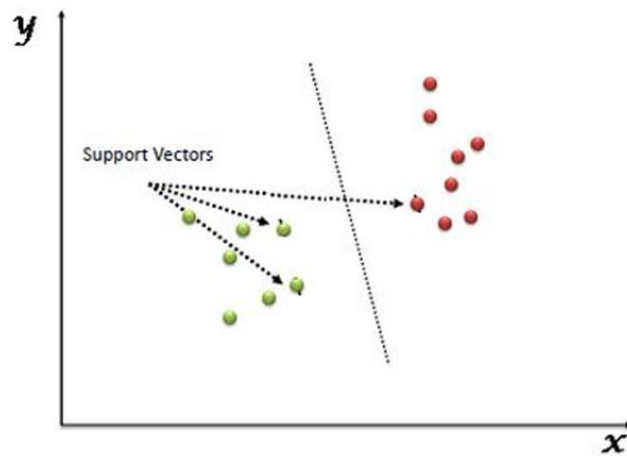
---

SVM é um algoritmo de aprendizagem de máquina supervisionado que classifica casos com base na determinação de um separador

- Mapeia-se os dados em uma dimensão maior para o espaço dos atributos
- Encontra-se então o separador

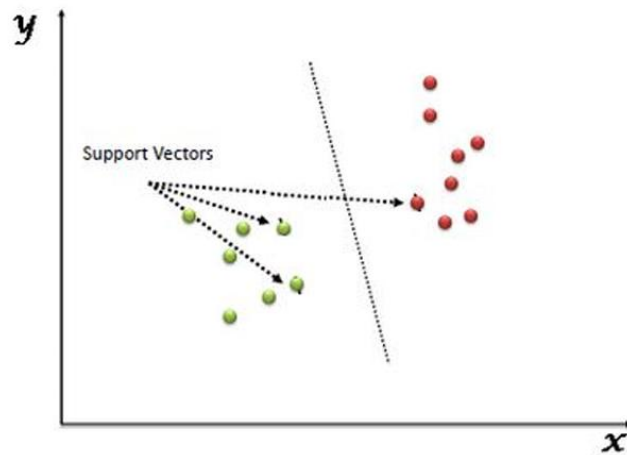
# SUPPORT VECTOR MACHINES (SVM)

Nesse algoritmo, plotamos cada item de dados como um ponto no espaço  $n$ -dimensional (onde  $n$  é o número de recursos que você tem), com o valor de cada recurso sendo o valor de uma determinada coordenada. Então, nós executamos a classificação encontrando o hiperplano que melhor diferencia



# SUPPORT VECTOR MACHINES (SVM)

Os “Vetores de suporte” são simplesmente as coordenadas da observação individual. Support Vector Machine é uma fronteira que melhor segrega as duas classes (hiperplano / linha).



# Classificação com SVM

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000015	6	1	1	1	7	1	3	1	1	Benign

Modeling

Prediction



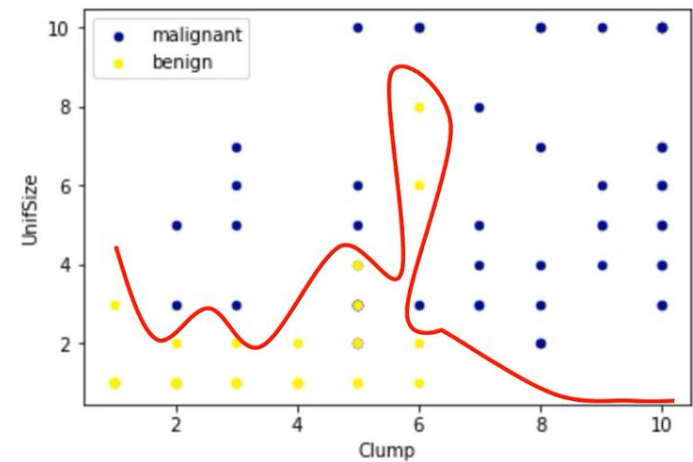
Accuracy = 89%

# Classificação com SVM

Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
5	1	1	1	2	1	3	1	1	benign
5	4	4	5	7	10	3	2	1	benign
3	1	1	1	2	2	3	1	1	malignant
6	8	8	1	3	4	3	7	1	benign
4	1	1	3	2	1	3	1	1	benign
8	10	10	8	7	10		7	1	malignant
1	1	1	1	2	10	3	1	1	benign
2	1	2	H	2	1	3	1	1	benign
2	1	1	1	2	1	1	1	5	benign
4	2	1	1	2	1	2	1	1	benign

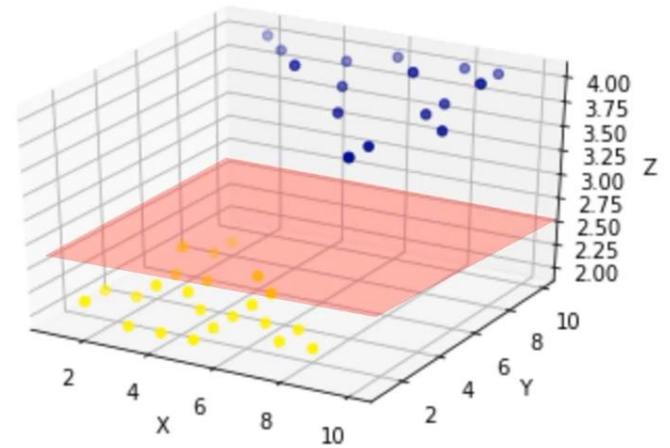
# Classificação com SVM

Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
5	1	1	1	2	1	3	1	1	benign
5	4	4	5	7	10	3	2	1	benign
3	1	1	1	2	2	3	1	1	malignant
6	8	8	1	3	4	3	7	1	benign
4	1	1	3	2	1	3	1	1	benign
8	10	10	8	7	10		7	1	malignant
1	1	1	1	2	10	3	1	1	benign
2	1	2	H	2	1	3	1	1	benign
2	1	1	1	2	1	1	1	5	benign
4	2	1	1	2	1	2	1	1	benign



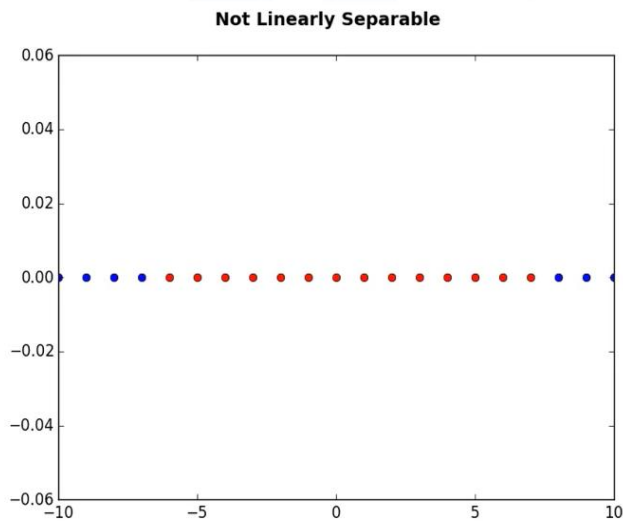
# Classificação com SVM

Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
5	1	1	1	2	1	3	1	1	benign
5	4	4	5	7	10	3	2	1	benign
3	1	1	1	2	2	3	1	1	malignant
6	8	8	1	3	4	3	7	1	benign
4	1	1	3	2	1	3	1	1	benign
8	10	10	8	7	10		7	1	malignant
1	1	1	1	2	10	3	1	1	benign
2	1	2	H	2	1	3	1	1	benign
2	1	1	1	2	1	1	1	5	benign
4	2	1	1	2	1	2	1	1	benign



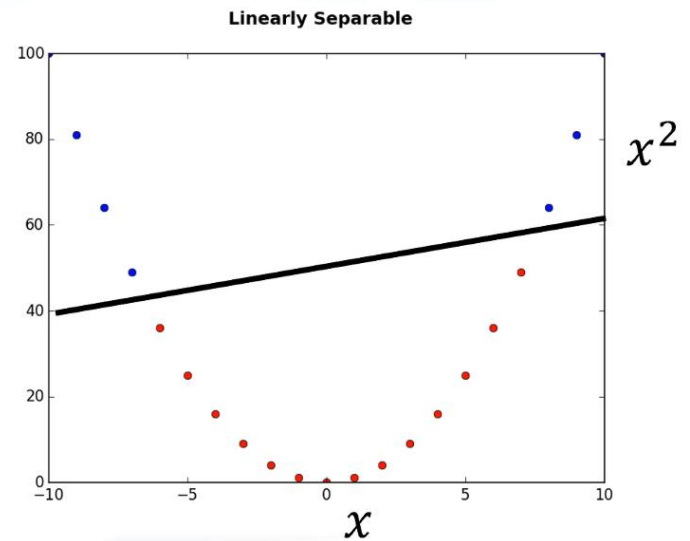
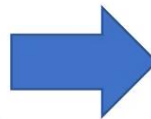


# Transformação dos dados



**Kernelling:**

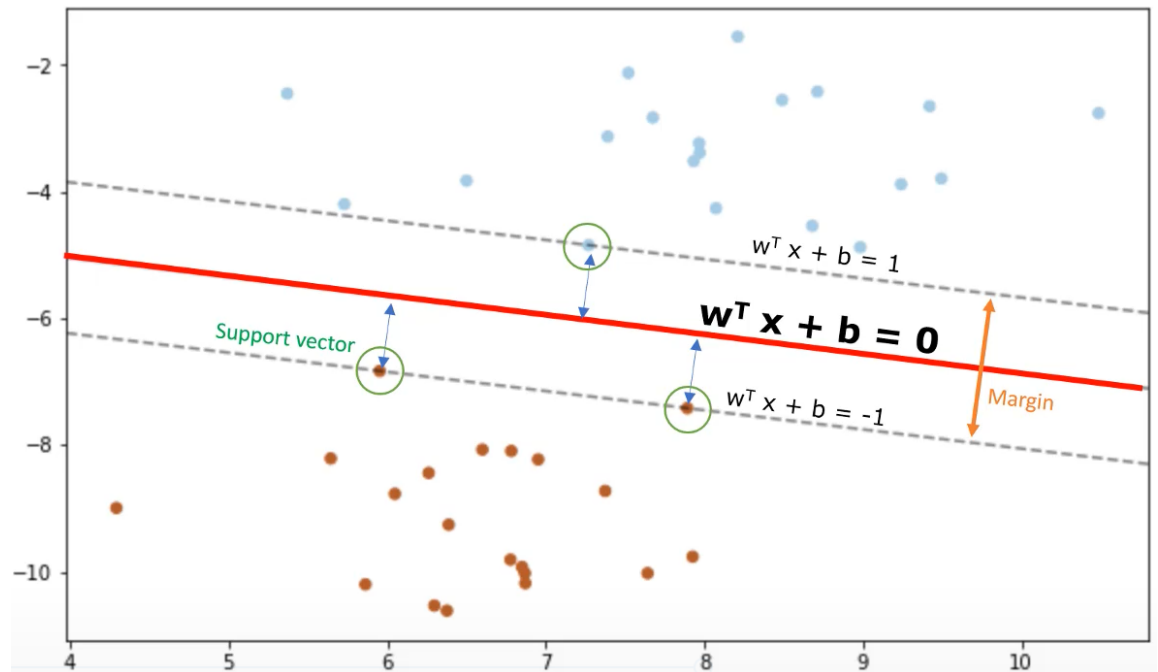
- Linear
- Polynomial
- RBF
- Sigmoid



$$\phi(x) = [x, x^2]$$

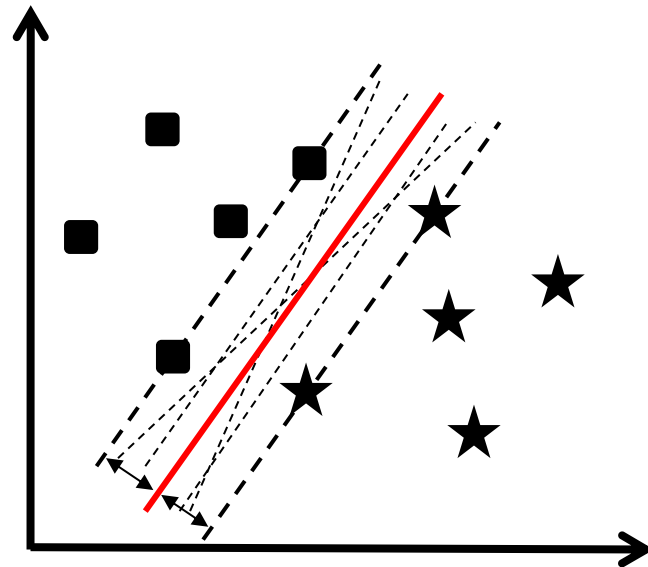
# Usando o SVM para encontrar o hiperplano

Find  $\mathbf{w}$  and  $b$  such that  
 $\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}$  is minimized;  
and for all  $\{(\mathbf{x}_i, y_i)\}$ :  $y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1$



# Support Vector Machine

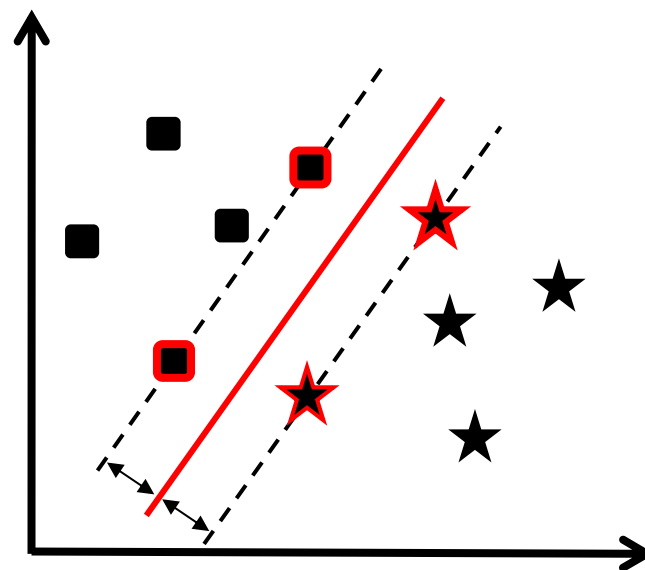
- Como separar essas duas classes?
  - Existem diversas retas que podem ser traçadas para separar os dados.
- Qual delas é a melhor opção?



**Hiperplano ótimo!**

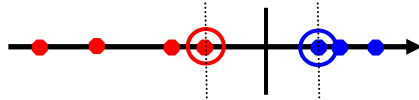
# Vetores de Suporte

- Servem para definir qual será o hiperplano.
- São encontrados durante a fase de treinamento.
- Os vetores de suporte são os exemplos de treinamento realmente importantes. Os outros exemplos podem ser ignorados.

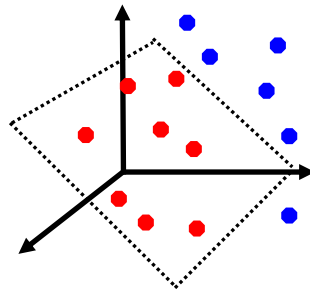


# Support Vector Machine

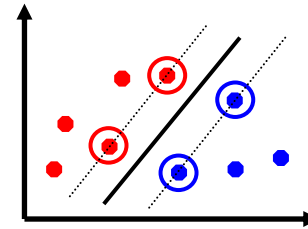
- Hiperplano:
  - Espaço 1D = Ponto



- Espaço 3D = Plano



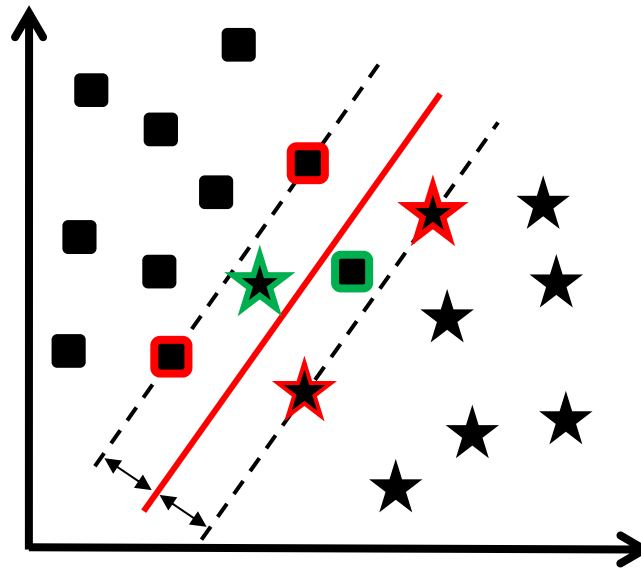
Espaço 2D = Reta



# Soft Margin

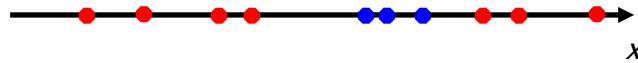
Soft Margin introduz uma variável de folga para medir o grau de erro de classificação.

Permite que alguns pontos sejam classificados de forma incorreta, desde que a maioria dos pontos seja classificada corretamente.

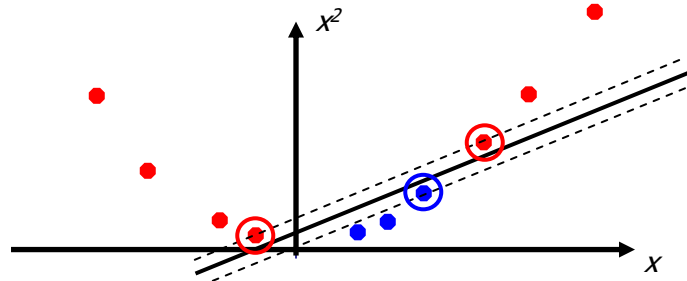


# SVM Não-Linear

O que fazer quando os dados não são linearmente separáveis?

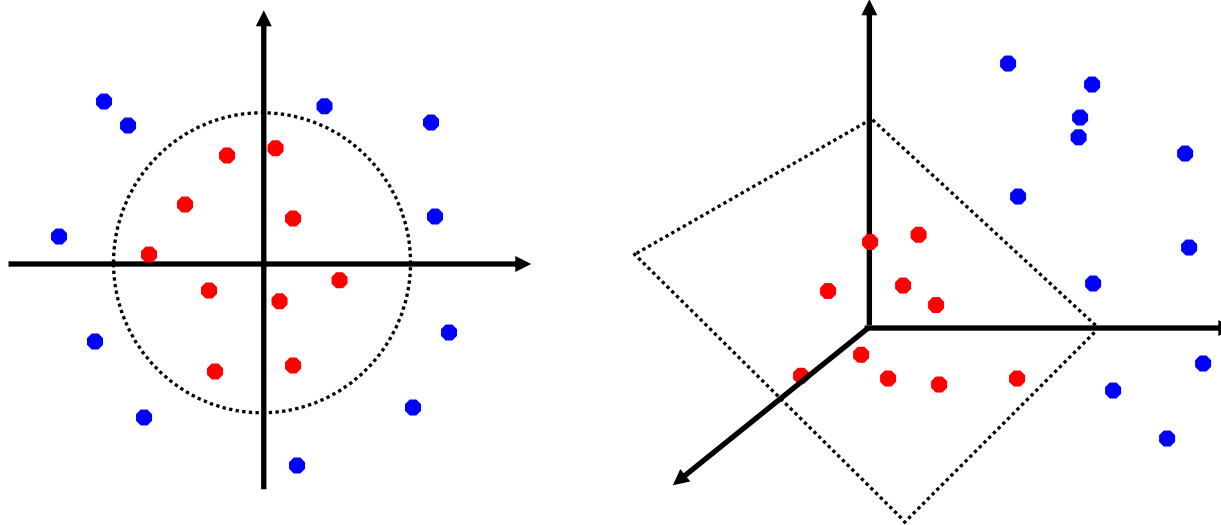


A abordagem utilizada pelo SVM para resolver esse tipo de problema consiste em mapear os dados para um espaço de dimensão maior:



# SVM Não-Linear

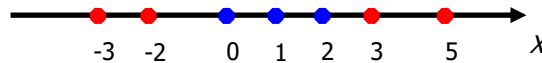
O espaço de atributos original pode ser mapeado em um espaço de atributos de dimensão maior onde o conjunto de treinamento é linearmente separável:



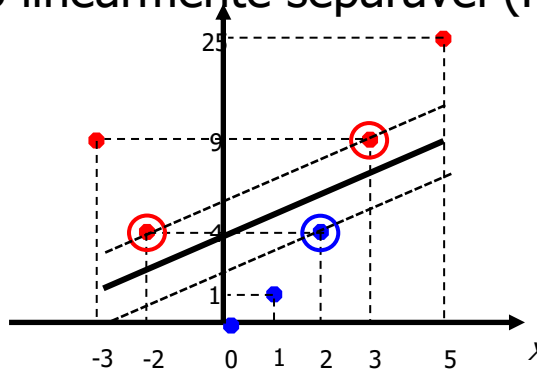


# SVM Não-Linear Exemplo

- Considerando o seguinte conjunto de exemplos de treinamento que não são linearmente separáveis:



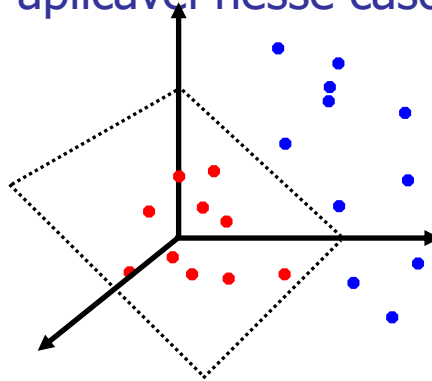
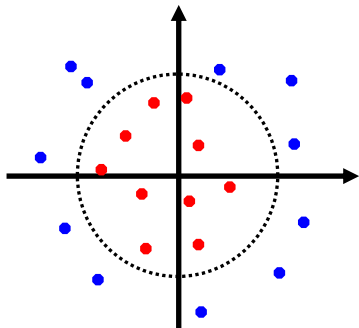
- Elevando para uma dimensão linearmente separável ( $\mathbb{R}^1 \rightarrow \mathbb{R}^2$ ):



- Kernel:  $\phi(x) = (x, x^2)$

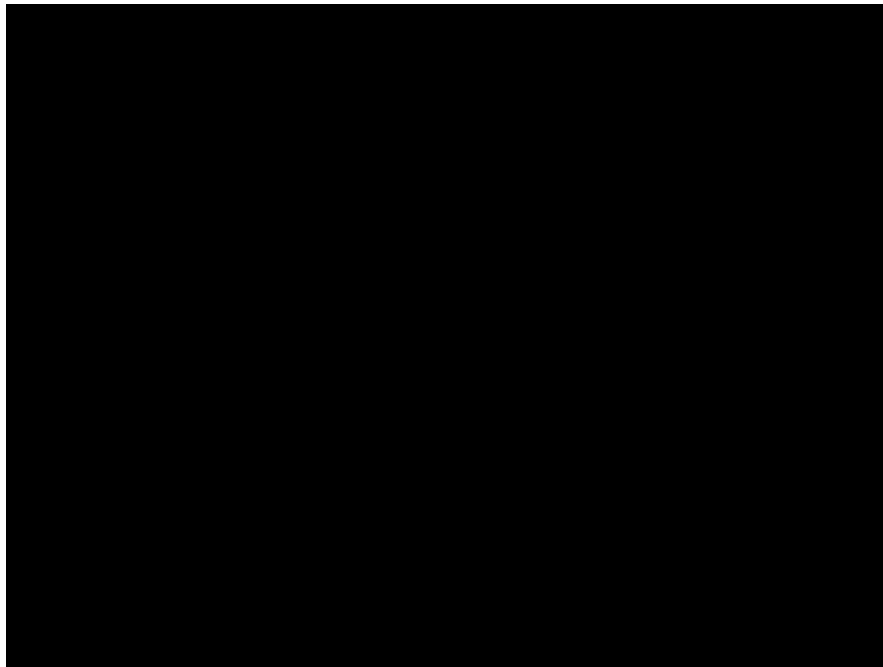
# SVM Não-Linear Exemplo

- A mesma metodologia pode ser aplicada em um espaço 2D de características ( $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ ).
- A única diferença é a necessidade de uma nova função de kernel. Um exemplo de função de kernel aplicável nesse caso seria:



# Polynomial Kernel

---



# Pontos Fortes e Fracos SVM

## PONTOS FORTES

- Preciso para classificação em espaços de múltiplas dimensões
- Eficiente em memória
- O processo de classificação é rápido.

## PONTOS FRACOS

- É necessário definir um bom Kernel.
- Tendencioso a overfitting se as features são maiores que as amostras
- Não dá estimativas de probabilidades
- Não eficiente em grandes datasets

Critério	Árvore de Decisão	Random Forest	Gradient Boosting (GBM)	Support Vector Machine (SVM)
Tipo	Modelo único de árvore (simples)	Conjunto de árvores (bagging)	Conjunto de árvores (boosting)	Classificador baseado em margens (hiperplanos)
Tendência ao overfitting	Alta (se não for podada)	Menor, pois combina várias árvores	Menor que árvore simples, mas pode superajustar	Pode overfit se C muito alto ou kernel mal escolhido
Interpretação	Muito interpretável	Pouco interpretável	Menos interpretável que RF	Difícil de interpretar
Velocidade de treino	Muito rápida	Razoável	Mais lenta que RF	Lenta em bases grandes
Velocidade de predição	Muito rápida	Rápida	Razoável	Razoável, mas pode ser lenta
Escalabilidade	Boa	Boa	Razoável	Ruim em datasets enormes

Critério	Árvore de Decisão	Random Forest	Gradient Boosting (GBM)	Support Vector Machine (SVM)
Escalabilidade	Boa	Boa	Razoável	Ruim em datasets enormes
Normalização necessária	Não	Não	Não	Sim
Capacidade p/ não-lineares	Fraca	Boa	Muito boa	Excelente com kernel
Importância das variáveis	Fácil de extrair	Fácil de extrair	Fácil de extrair	Não nativa (usar permutação)
Aplicação típica	Modelos simples e interpretáveis	Classificação e regressão em grandes bases	Alta performance em competições	Dados complexos, NLP, bioinformática
Vantagens	Fácil de entender e visualizar	Reduz overfitting, robusto	Alta acurácia, flexível	Bom em alta dimensão
Desvantagens	Overfitting, baixa precisão sozinho	Menos interpretável, mais lento	Mais lento e sensível a hiperparâmetros	Difícil de escalar, ajuste fino

Critérios de Comparação			
Critério	Árvore de Decisão	Random Forest	Gradient Boosting (GBM)
Tipo	Modelo único de árvore (simples)	Conjunto de árvores (bagging)	Conjunto de árvores (boosting)
Tendência ao overfitting	Alta (se não for podada)	Menor, pois combina várias árvores	Menor que árvore simples, mas pode superajustar
Interpretação	Muito interpretável	Pouco interpretável	Menos interpretável que RF
Velocidade de treino	Muito rápida	Razoável	Mais lenta que RF
Velocidade de predição	Muito rápida	Rápida	Razoável
Escalabilidade	Boa	Boa	Razoável
Normalização necessária	Não	Não	Não