

Machine Learning com Python

Prof. Luciano Galdino

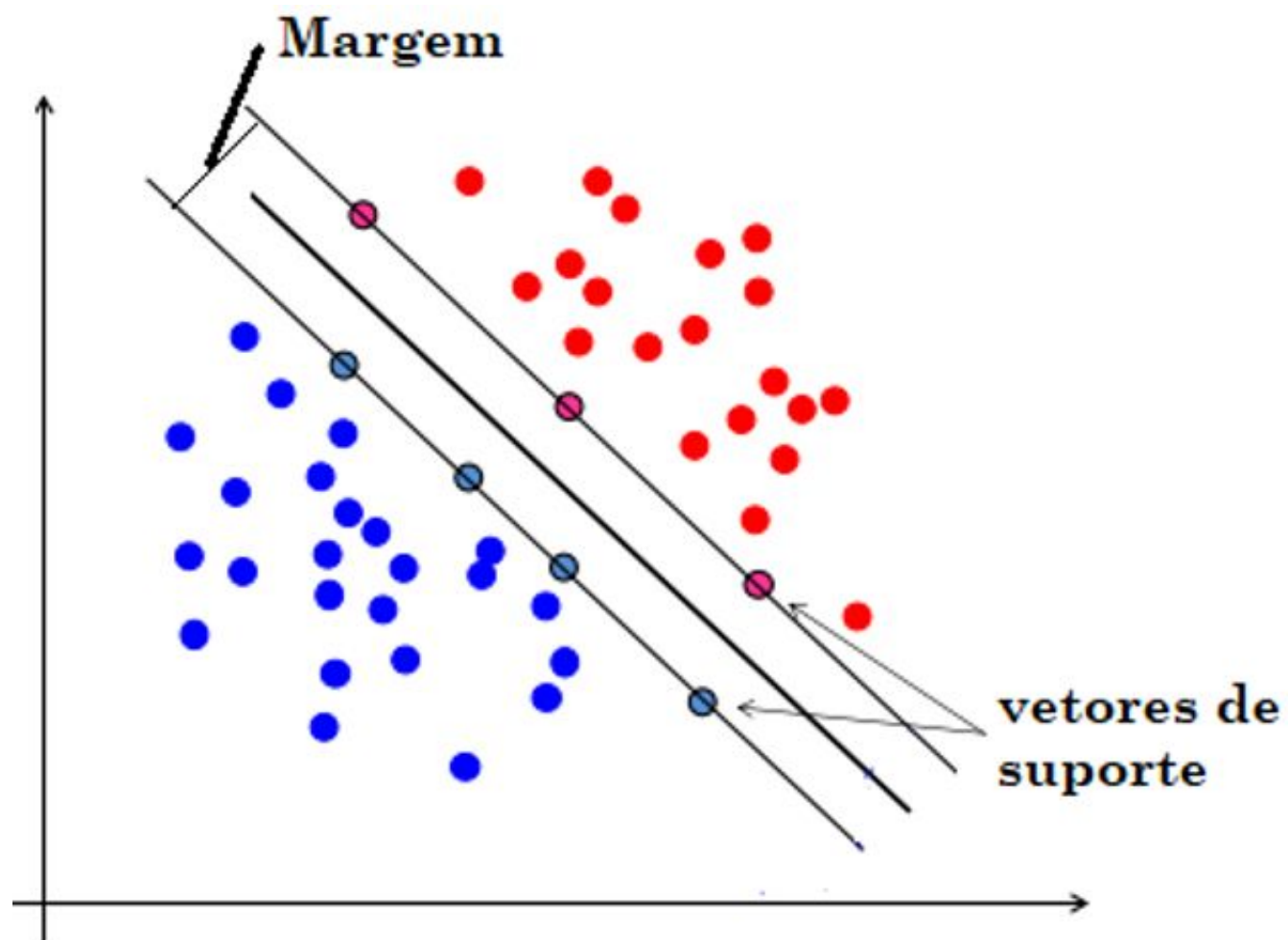
MÁQUINAS DE VETORES DE SUPORTE (SVM)

Aplicado em problemas de aprendizagem supervisionada tanto de classificação como de regressão.

Em classificação é conhecido como Classificador de Vetor de Suporte (SVC).

Cria hiperplanos de separação.

Pode ser aplicado em problemas linearmente separáveis e não linearmente separáveis.



Equação

$$\vec{w} \cdot \vec{x} + b \geq 0$$

Onde:

w = vetor hiperplano

x = vetor pertencente ao hiperplano.

b = deslocamento em relação à origem.

Algumas Aplicações

Classificação.

Categorização de textos.

Reconhecimento de imagem.

Detecção facial.

Detecção de anomalias.

Reconhecimento de letras manuscritas.

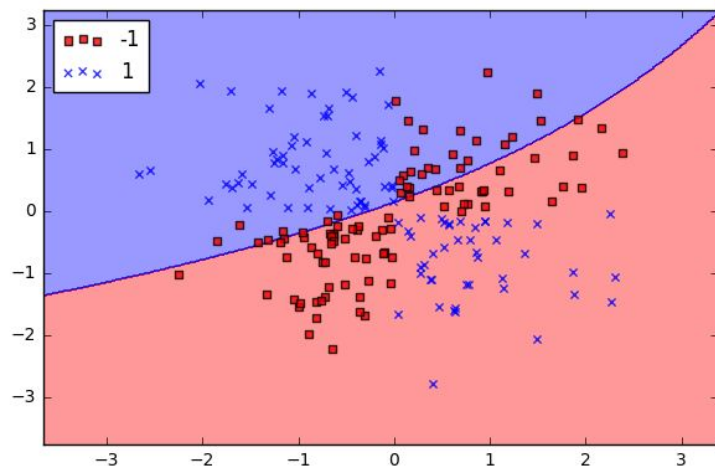
Constante de penalização (custo)

Hiperparâmetro C: controla a tolerância dos erros.

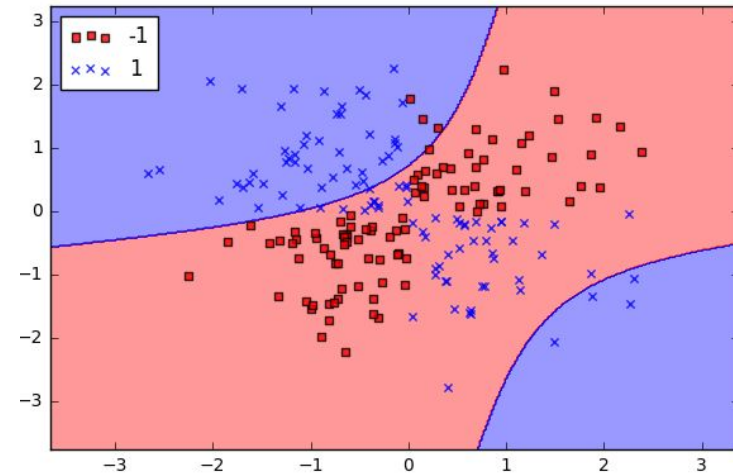
Quanto maior o valor de C, maior o poder de separação das classes, porém maior a probabilidade de overfitting e do tempo de treinamento.

Quanto menor o valor C, maior a chance de erros na separação e conseqüentemente ocasiona underfitting.

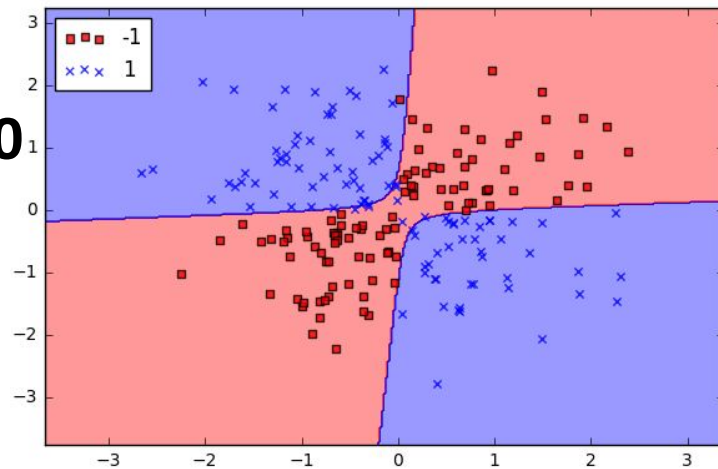
C=10



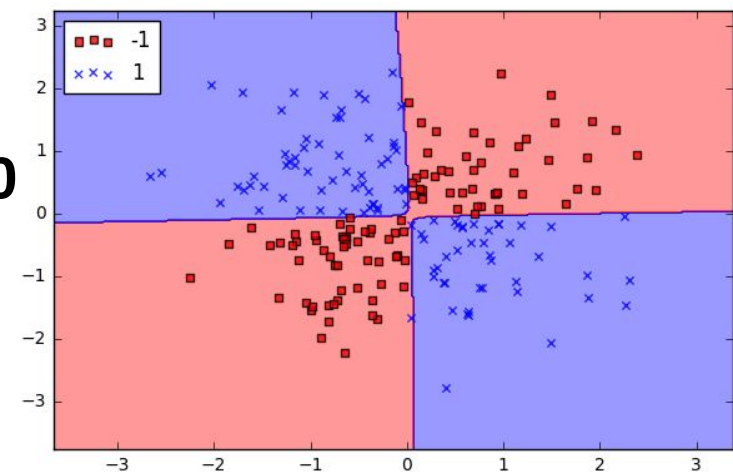
C=100



C=1000

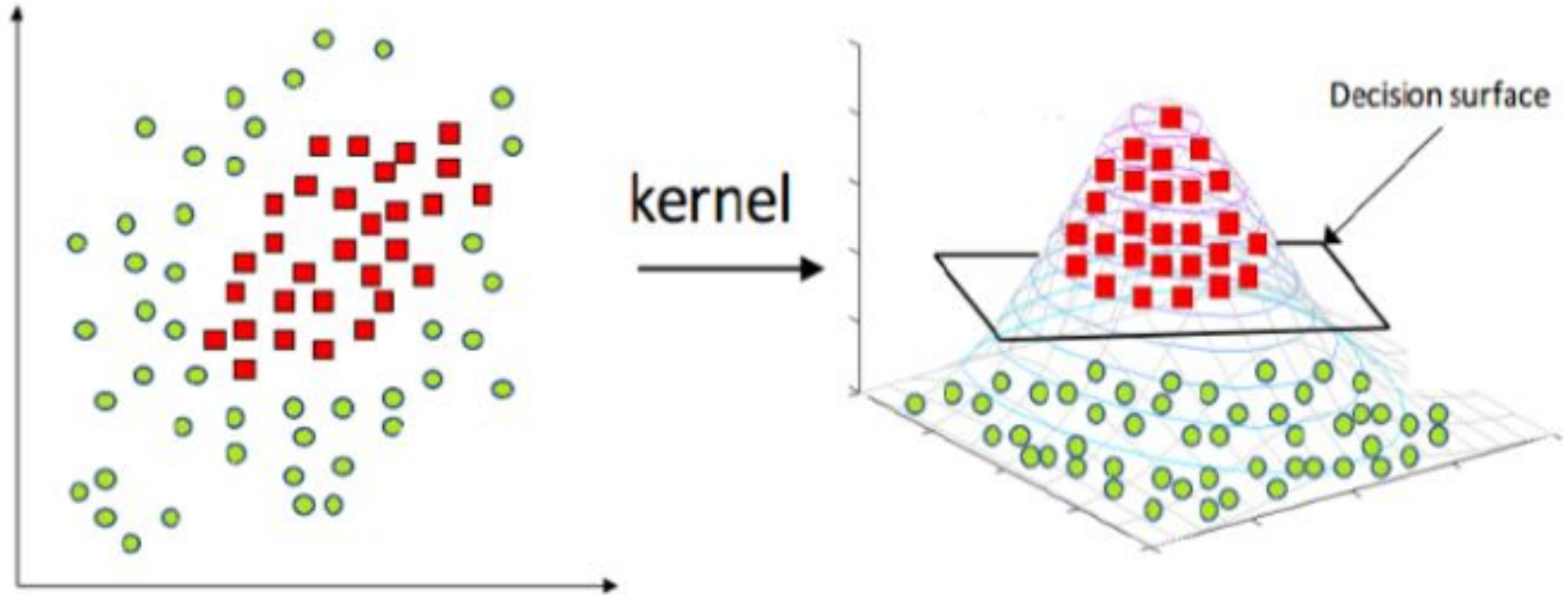


C=10000

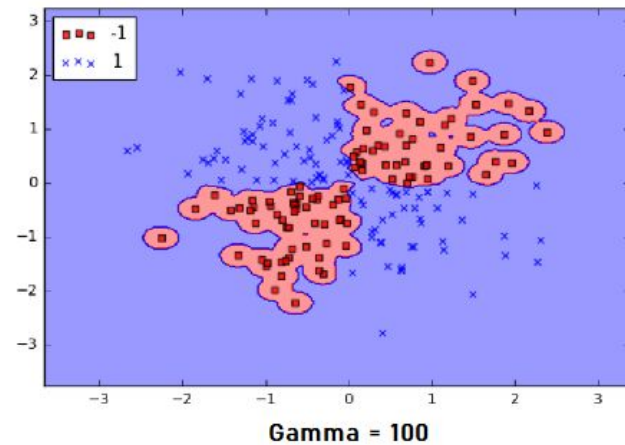
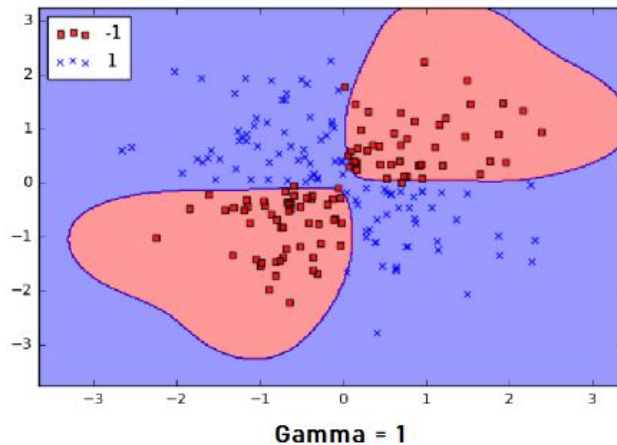
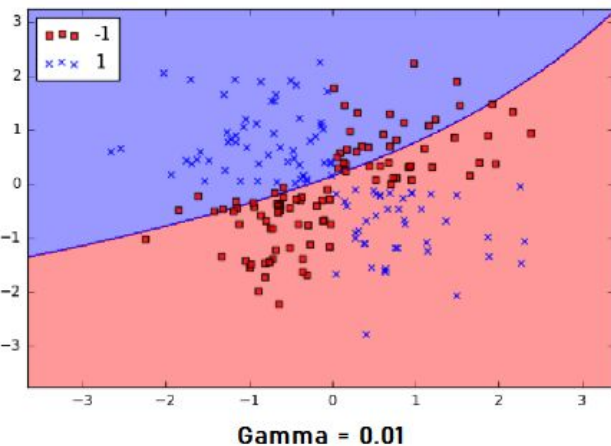


Problemas não lineares

Utiliza-se a técnica kernel Trick para transformação de não lineares para lineares.



Ajusta-se o hiperparâmetro γ para otimização.



https://chrisalbon.com/code/machine_learning/support_vector_machines/svc_parameters_using_rbf_kernel/

Vantagens

Não é influenciado por dados discrepantes.

Solução de problemas lineares e não lineares.

Muito efetivo para datasets grandes.

Consegue aprender com características não pertencentes aos dados.

Desvantagens

Difícil interpretação teórica devido a matemática complexa.

Difícil visualização gráfica.

É lento comparado aos outros algoritmos.

Deve-se ter um grande cuidado com as definições dos hiperparâmetros para evitar overfitting e underfitting.