



Teoria do Aprendizado Estatístico

Atividade 4

Luiz Henrique Barretta Francisco - 202100155302

maio/2025

Introdução

Vários problemas de classificação envolvem dados que não são linearmente separáveis. Nesses cenários, classificadores lineares como o SVM simples, tanto com margens rígidas ou suaves, falham em encontrar um hiperplano que separe as classes de forma satisfatória, considerando o espaço original dos dados. Para resolver isso, o *Support Vector Machine* (SVM) utilizado o chamado ‘truque do kernel’.

A principal ideia do método é mapear os dados de entrada em uma dimensão menor para um espaço de uma dimensão muito maior, onde lá os dados ficam linearmente separáveis. Entretanto, a função de mapeamento que liga essas dimensões pode ser computacionalmente caro ou até inviável. O ‘truque do kernel’ serve para que o SVM opere nesse espaço de dimensão maior de forma implícita, assim calculando os produtos escalares entre as imagens (fruto de uma transformação linear T) dos dados nesse espaço sem precisar calcular esse mapeamento em si.

Um dos kernels mais populares e eficazes para lidar com relações que não são lineares nos dados é o kernel Gaussiano de Função de Base Radial (RBF), que iremos definir e simular nesse trabalho.

Metodologia

O kernel Gaussiano, ou RBF, é definido por:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

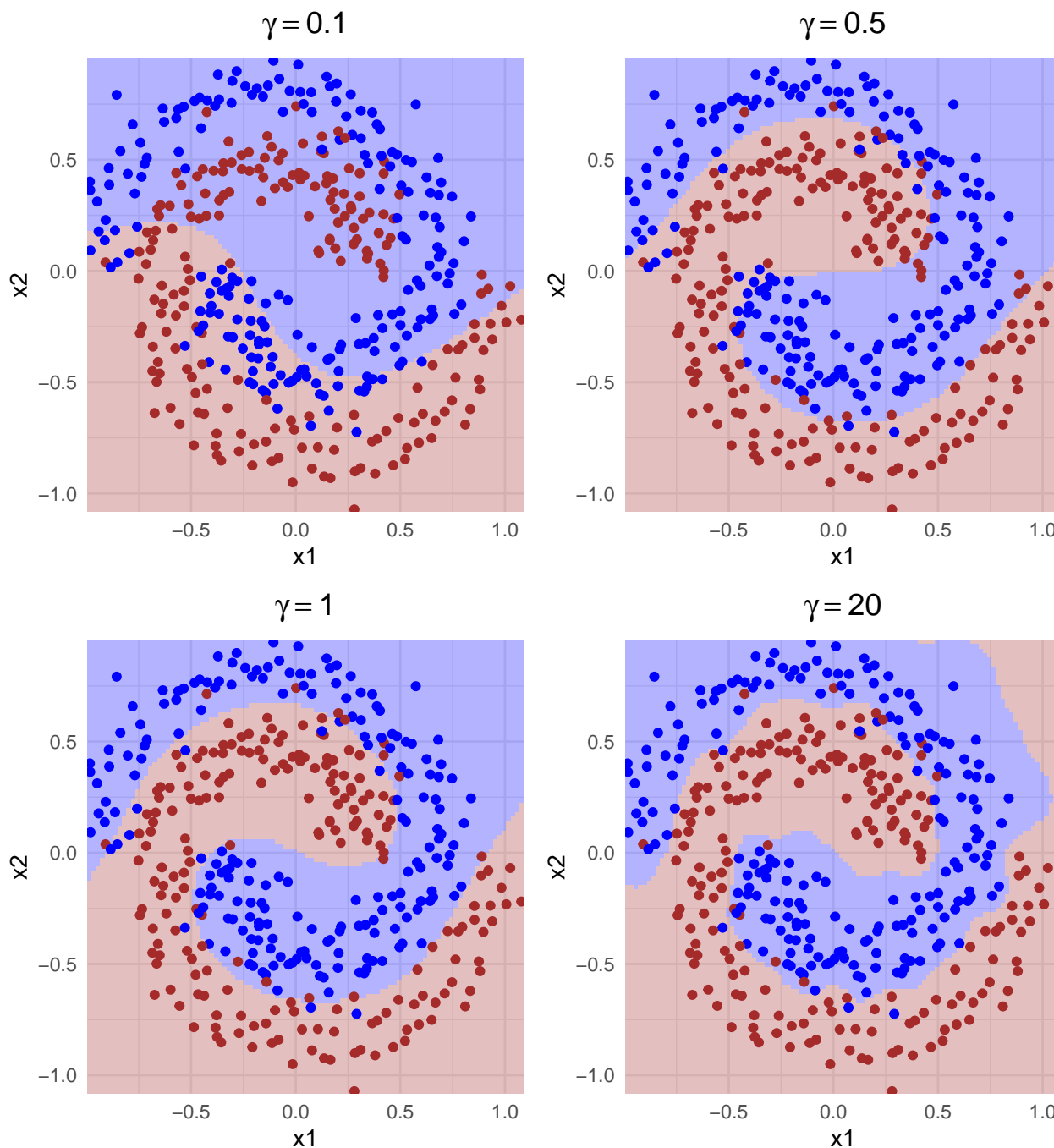
com:

$$\gamma = \frac{1}{2\sigma^2}, \quad \gamma > 0$$

onde: x_i e x_j são os dois vetores de observações. $\|x_i - x_j\|^2$ é o quadrado da distância Euclidiana entre os pontos. γ é um hiperparâmetro que irá ditar a ‘largura’ do kernel. Um valor baixo para γ significa que a influência de um único ponto de treinamento alcança uma longa distância, resultando em um limite de decisão com margens mais suaves. Um valor alto de γ significa que a influência é curta, causando um limite de decisão mais complexo com margens rígidas e possivelmente gerando um *overfitting* aos dados de treinamento. A função de mapeamento $\phi(x)$ mapeia os dados para um espaço de dimensão infinita. A vantagem do truque do kernel é que não precisamos calcular $\phi(x)$ de forma explícita. O algoritmo SVM precisa apenas dos valores de $K(x_i, x_j)$, que medem a similaridade entre x_i e x_j . Se dois pontos são próximos no espaço original, o valor do kernel é próximo de 1. Se são distantes, o valor do kernel é próximo de 0, gerando assim a divisão entre classes.

Para ilustrar a eficácia do método do kernel Gaussiano RBF, utilizaremos um conjunto de dados bidimensional com duas classes dispostas em uma espiral. Esse conjunto de dados é um exemplo clássico de classificação não-linear.

Efeito do Hiperparâmetro Gamma na Fronteira de Decisão do SVM

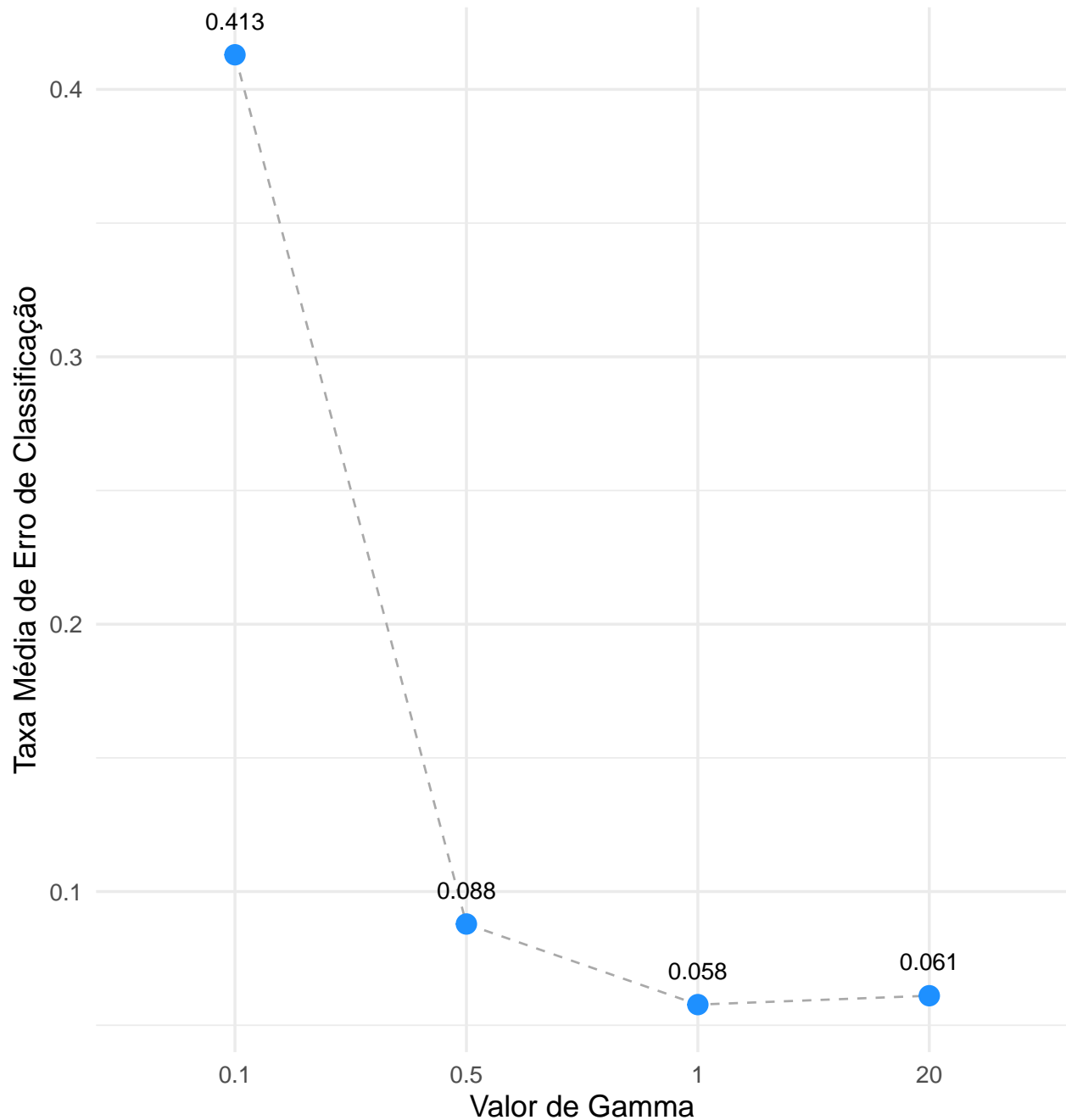


Diferentes fronteiras de decisão.

Os gráficos acima ilustram como o hiperparâmetro γ afeta a fronteira de decisão do modelo SVM com kernel RBF. Com um valor baixo, a fronteira é muito suave e simples, falhando em separar as duas classes (causando subajuste). À medida que γ aumenta, a fronteira se torna mais flexível e se adapta melhor à forma não-linear dos dados, conseguindo uma separação adequada quando $\gamma = 1$. No entanto, com um valor muito alto, a fronteira se torna excessivamente complexa e irregular, ajustando-se até mesmo a pontos individuais e irrelevantes no geral. Isso indica um superajuste (overfitting), onde o modelo perde a sua capacidade de generalização.

Erro Médio de Generalização vs. Hiperparâmetro Gamma

Estimado a partir de 1000 simulações com novos dados



O gráfico do erro de generalização corrobora a suposição de overfitting levantada anteriormente. Ele confirma que um $\gamma = 0.1$ muito baixo resulta em um modelo com alto erro (41.3%), por ser simples demais e não aprender o padrão dos dados. À medida que γ aumenta, o erro cai drasticamente, atingindo seu ponto mais baixo em 5.8% para $\gamma = 1$, indicando o melhor equilíbrio entre simplicidade e complexidade. O aumento do erro com $\gamma = 20$ para 6.1% é um indicativo de que o modelo pode estar começando a se ajustar demais aos dados de treinamento e tendo um overfitting, o que prejudica sua performance em dados novos e nunca vistos.

Comentários Finais

Como mostrado, o uso de funções de mapeamento, especialmente ao utilizar o truque do kernel, confere ao SVM uma grande flexibilidade para resolver problemas de classificação mais complexos. A capacidade do kernel Gaussiano (RBF) de separar dados não lineares se baseia no princípio de transformar o espaço de variáveis original em um espaço de dimensão muito maior (teoricamente infinita), onde os dados se tornam linearmente separáveis.

O truque do kernel permite que essa operação seja feita de forma computacionalmente eficiente. Em vez de calcular as novas coordenadas de cada ponto de dado nesse espaço de alta dimensão, o kernel RBF calcula diretamente uma medida de similaridade entre os pontos no espaço original. A fórmula $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ atribui um valor alto a pontos que estão próximos e um valor baixo a pontos distantes. O SVM utiliza essas medidas de similaridade para construir uma fronteira de decisão que, embora seja um simples hiperplano no espaço de alta dimensão, se manifesta como uma fronteira complexa e não-linear quando projetada de volta ao espaço original dos dados. É por isso que ele consegue “desenhar” curvas, círculos ou formas mais complexas para separar as classes.

O hiperparâmetro γ é crucial nesse processo, pois é ele quem determina essa proximidade. Um valor de γ alto considera apenas pontos muito próximos como similares, levando a uma fronteira mais complexa e ajustada (causando um risco de overfitting). Um γ baixo tem uma visão de similaridade mais ampla, resultando numa fronteira mais suave (agora com risco de underfitting). Esse ajuste fino do hiperparâmetro é um exemplo clássico do compromisso entre viés e variância, onde o objetivo final é minimizar o erro de generalização do modelo. A escolha ótima de γ depende fundamentalmente da estrutura dos dados, não existindo um valor universal que sirva para todos os problemas.

Referências

ARA, Anderson; OSPINA, Raydonal; MAIA, Mateus. **Modelos de Vetores de Suporte: Uma Introdução ao Aprendizado Estatístico de Máquina**. 2023. Material apresentado na 67ª RBRas e 20º SEAGRO, Londrina-PR. Disponível em: <http://leg.ufpr.br/~ara/teach/svm/>. Acesso em: 9 jun. 2025.

CORTES, Corinna; VAPNIK, Vladimir. Support-Vector Networks. **Machine Learning**, v. 20, n. 3, p. 273-297, 1995.

LEISCH, Friedrich; DIMITRIADOU, Evgenia. **mlbench: Machine Learning Benchmark Problems**. R package version 2.1-3, 2021. Disponível em: <https://CRAN.R-project.org/package=mlbench>.