Support Vector Machines (SVM)

O algoritmo Support Vector Machine (SVM) constitui uma das abordagens mais consolidadas no contexto do aprendizado supervisionado, sendo eficaz na resolução de problemas de classificação binária. Sua fundamentação repousa em princípios geométricos e de otimização convexa, cujo objetivo central é encontrar a hipersuperfície ótima que separa dois conjuntos de dados pertencentes a classes distintas, maximizando a margem entre tais classes no espaço de atributos.

Formalmente, um SVM busca identificar um hiperplano de separação $w \cdot x + b = 0$ tal que os dados de treinamento sejam corretamente classificados e a distância entre o hiperplano e os exemplos mais próximos de cada classe — denominados vetores de suporte — seja maximizada. A maximização dessa margem constitui o critério de generalização adotado pelo algoritmo, uma vez que uma separação com margem ampla tende a ser mais robusta à presença de ruídos nos dados e a variações futuras no conjunto de entrada.

O que é um hiperplano de separação?

Imagine que você tem duas espécies de frutas espalhadas em uma mesa: maçãs e laranjas. Cada fruta tem duas características: o peso e a cor. Se você fizer um gráfico com essas duas informações, cada fruta vira um ponto no plano.

Agora, pense em desenhar uma reta que consiga separar as maçãs das laranjas — deixando as maçãs de um lado e as laranjas do outro. Essa reta é o que chamamos, nesse caso, de hiperplano de separação (como estamos no plano, o “hiperplano” é apenas uma reta; em 3D seria um plano, e em espaços com mais características, uma superfície mais complexa).

A equação do hiperplano: $w \cdot x + b = 0$

- x representa os dados da fruta (por exemplo, seu peso e sua cor, em forma de números).
- w é um vetor de pesos que diz quanto cada característica importa para separar as frutas.
- . representa o produto escalar, que basicamente junta os dados com os pesos.
- b é um número que ajusta a posição da reta no gráfico.
- E tudo isso igual a zero (= 0) define exatamente onde a reta passa.

O que essa equação faz é traçar uma linha imaginária que divide o espaço em dois lados:

- De um lado, estão os pontos que resultam em $w \cdot x + b > 0 \rightarrow$ por exemplo, as maçãs.
- Do outro lado, estão os pontos com $w \cdot x + b < 0 \rightarrow$ por exemplo, as laranjas.
- Os pontos sobre a linha são os que resultam em $w \cdot x + b = 0$.

Então, identificar um hiperplano significa descobrir quais números de w e b devemos usar para que a equação $w \cdot x + b = 0$ desenhe a melhor linha possível para separar os dois grupos.

É como programar um robô para olhar um monte de frutas e traçar, sozinho, uma linha que diga: "Tudo que estiver deste lado é maçã, e tudo que estiver do outro é laranja."

Nos casos em que os dados são linearmente separáveis, a formulação primal do problema de otimização consiste em minimizar $\frac{1}{2}|w|$ sujeito às restrições $y_i(w \cdot x_i + b) \geq 1$, onde $y_i \in -1, +1$ representa a classe do exemplo x_i. Em outras palavras, o SVM procura a linha que separa bem os dados, ficando o mais distante possível dos pontos. Ele faz isso escolhendo os pesos adequados w para que cada ponto fique do lado correto, garantindo uma margem ampla. Dessa forma, a solução ótima depende apenas de um subconjunto dos exemplos de treinamento: os vetores de suporte. São eles os únicos pontos que satisfazem a condição de igualdade nas restrições e, portanto, definem geometricamente o posicionamento do hiperplano.

Contudo, a suposição de separabilidade linear nem sempre é válida em contextos reais. Para contornar essa limitação, o SVM incorpora o chamado método do kernel, que viabiliza a separação não linear por meio de uma transformação implícita dos dados para um espaço de dimensão superior, no qual a separabilidade linear pode ser recuperada. Essa técnica, conhecida como "*kernel trick*", permite que o produto interno entre vetores transformados seja calculado diretamente por meio de uma função núcleo $K(x_i, x_j)$, sem a necessidade de explicitar o espaço de mapeamento.

As funções kernel mais utilizadas incluem o kernel polinomial, o kernel sigmoide e, principalmente, o kernel radial de base (RBF). Este último é eficaz em problemas onde a fronteira de decisão é altamente não linear, pois projeta os dados em um espaço de características com estrutura de similaridade centrada em cada ponto. A escolha do kernel e de seus parâmetros associados (como o grau no polinomial ou o parâmetro γ no RBF) exerce influência direta na capacidade do modelo de capturar padrões relevantes, mas também na propensão ao sobreajuste, sendo, portanto, objeto de validação cruzada criteriosa.

Além de sua aplicação na classificação binária, o SVM pode ser estendido a contextos de classificação multiclasse por meio de estratégias como *one-vs-rest* ou *one-vs-one*, bem como ser adaptado para regressão (SVR — Support Vector Regression), mantendo o mesmo princípio de margem, mas em torno de uma função de predição contínua.

Em síntese, o SVM destaca-se por combinar uma formulação matemática rigorosa com propriedades desejáveis do ponto de vista computacional e estatístico, como a convexidade do problema de otimização (que assegura a unicidade da solução) e a robustez frente a conjuntos de dados de alta dimensionalidade. Tais características o tornam adequado para aplicações em reconhecimento de padrões, bioinformática, análise de textos, diagnóstico médico, entre outras áreas que demandam classificadores estáveis e interpretáveis.

Referência

LENZ, MAIKON L.; NEUMANN, FABIANO B.; SANTARELLI, RODRIGO; ET AL. FUNDAMENTOS DE APRENDIZAGEM DE MÁQUINA. PORTO ALEGRE: SAGAH, 2020. E-BOOK. P.235. ISBN 9786556900902. DISPONÍVEL EM: HTTPS://APP.MINHABIBLIOTECA.COM.BR/READER/BOOKS/9786556900902/.