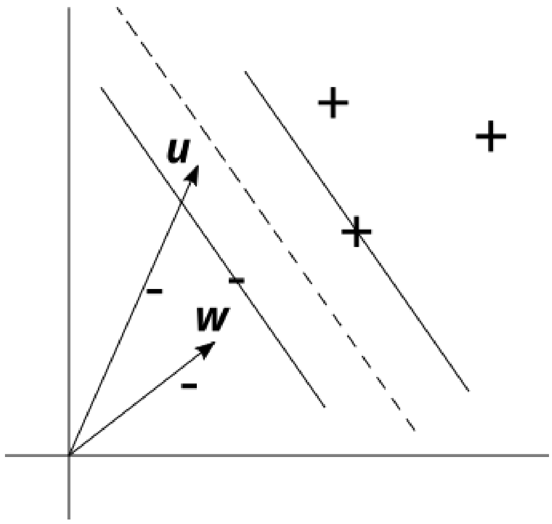
Support Vector Machine (SVM)
Autor: Gabriel Costa Leite

Support Vector Machines (SVMs) (Máquinas de vetores de suporte) é um conjunto de métodos de aprendizagem supervisionada utilizada na classificação, regressão e detecção de anomalias em um conjunto de dados.
A ideia principal do método é achar um divisor que tenha a melhor lei de decisão para classificar o conjunto de dados em estudo. Esse divisor é chamada do hiperplano. Tome então os conjuntos dos pontos + e −, como mostra a figura abaixo:



Sendo o vetor $\overrightarrow{w}$ normal ao hiperplano escolhido, podemos achar a componente do ponto $\overrightarrow{u}$ na direção normal e afirmar que se essa componente for maior a uma certo tamanho c o ponto será classificado como +. Logo, temos que:

$$\overrightarrow{w} \cdot \overrightarrow{u} \geq c$$

Sem perda de generalidade ($c = -b$):

$$\overrightarrow{w} \cdot \overrightarrow{u} + b \geq 0 \tag{1}$$

Ainda assim, iremos restringir que:

$$\begin{cases} \overrightarrow{w} \cdot \overrightarrow{x_+} + b \geq 1 \\ \overrightarrow{w} \cdot \overrightarrow{x_-} + b \leq -1 \end{cases} \tag{2}$$

Por conveniência matemática, podemos definir:

$$y_i = \begin{cases} +1 \rightarrow \text{ amostra } + \\ -1 \rightarrow \text{ amostra } - \end{cases} \tag{3}$$

Logo, multiplicando as equações da 2 respectivamente por y_i e $-y_i$, teremos que:

$$\begin{cases} y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geq 1 \\ y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) \geq 1 \end{cases} \tag{4}$$

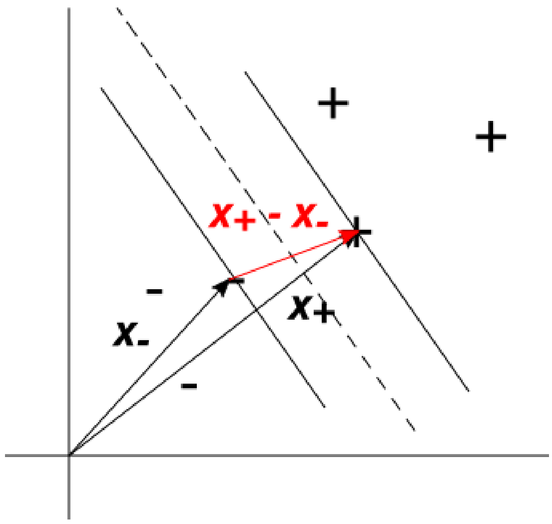
Observe que as equações são iguais e enfim chegamos a equação que define a restrição do modelo:

$$y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1 \geq 0 \tag{5}$$

Restringiremos ainda mais por conveniência, obtendo:

$$y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1 = 0 \tag{6}$$

O melhor divisor de dados será aquele que maximizará a distância entre a amostra $+$ e amostra $-$ mais próximo do divisor, como mostra a figura abaixo:



Vamos então maximizar essa distância que será igual a:

$$d = (\overrightarrow{x_+} - \overrightarrow{x_-}) \cdot \frac{\overrightarrow{w}}{||w||} \tag{7}$$

Substituindo 6 em 7:

$$d = (1 - b - (1 + b)) \cdot \frac{\overrightarrow{w}}{||w||} = \frac{2}{||w||} \tag{8}$$

Logo a ideia é achar o valor máximo para d, ou seja,

$$max(\frac{2}{||w||}) \therefore max(\frac{1}{||w||}) \therefore min(||w||) \therefore min(\frac{1}{2}||w||^2) \tag{9}$$

Então, o nosso problema é o mesmo de encontrar o valor mínimo da equação $\frac{1}{2}||w||^2$
Com efeito, ao optimizar essa equação, devemos honrar o as restrições definidas. Para encontrar o extremo de uma função com restrições devemos usar o multiplicadores de Lagrange, que nós dará a seguinte equação:

$$L = \frac{1}{2}||w||^2 - \sum \alpha_i[y_i(\overrightarrow{w} \cdot \overrightarrow{x_i} + b) - 1] \tag{10}$$

Para optimizar essa equação devemos derivar parcialmente e igualar a 0 L em relação a $\overrightarrow{w}$ e b

$$\begin{aligned} \frac{\partial L}{\partial \overrightarrow{w}} &= \overrightarrow{w} - \sum \alpha_i y_i \overrightarrow{x_i} = 0 \therefore \overrightarrow{w} = \sum \alpha_i y_i \overrightarrow{x_i} \\ \frac{\partial L}{\partial \overrightarrow{d}} &= \sum \alpha_i y_i = 0 \end{aligned} \tag{11}$$

Substituindo 11 em 10:

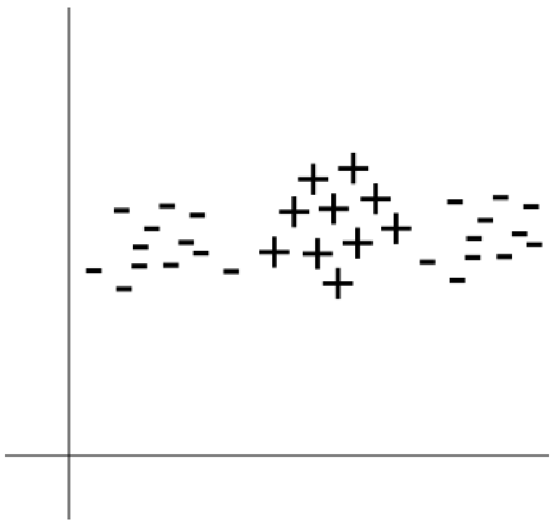
$$L = \frac{1}{2}(\sum \alpha_i y_i \overrightarrow{x_i}) \cdot (\sum \alpha_j y_j \overrightarrow{x_j}) - (\sum \alpha_i y_i \overrightarrow{x_i}) \cdot (\sum \alpha_j y_j \overrightarrow{x_j}) - \sum \alpha_i y_i b + \sum \alpha_i$$

$$L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \overrightarrow{x_i} \cdot \overrightarrow{x_j} \tag{12}$$

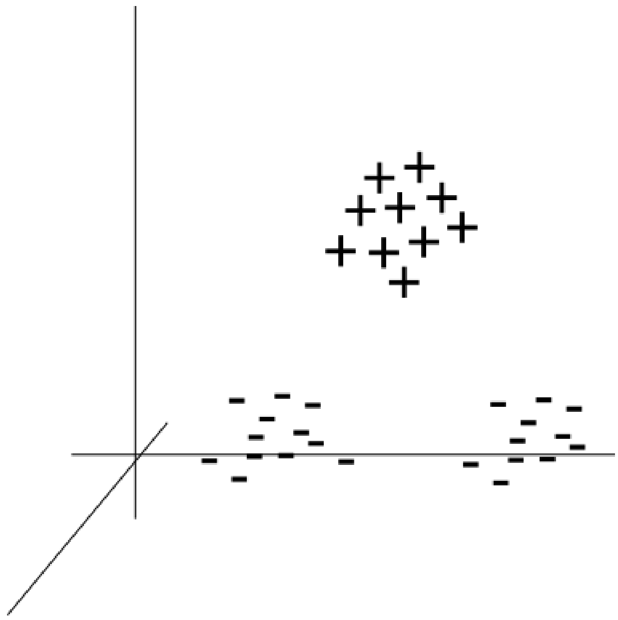
Logo, a nossa equação classificadora será:

$$\sum \alpha_i y_i \overrightarrow{x_i} \cdot \overrightarrow{x_j} + b \geq 0 \rightarrow \text{ amostra } + \tag{13}$$

Esse classificador funciona para certas amostras de dados, mas e se tivermos uma amostra como:



Não é possível encontrar um divisor adequado para esse conjunto. Para solucionar isso, é conveniente aumentar a dimensão em que se encontra a amostra, obtendo por exemplo isso:



Esse tipo de transformação requer um esforço computacional muito grande, mas se observarmos a equação 13, o que precisamos é um produto escalar. Se chamarmos ϕ a equação de transformação teremos que:

$$\sum \alpha_i y_i \overrightarrow{\phi(x_i)} \cdot \overrightarrow{\phi(x_j)} + b \geq 0 \rightarrow \text{ amostra } + \tag{14}$$

Assim podemos chamar $K(x_i, x_j) = \overrightarrow{\phi(x_i)} \cdot \overrightarrow{\phi(x_j)}$. Essa função é chamada de kernel e solução é chamada de kernel trick.
Uma das funções kernel mais utilizadas em problemas de SVM é a Radial Basis Function (RBF) ou função de base radial, onde:

$$K = e^{-\gamma(x_i - x_j)^2} \tag{15}$$

Esse tipo de kernel leva a nossa amostra para uma dimensão infinita e é possível provar isso fazendo a expansão da série de Taylor.
Dessa forma chegamos a nossa função de decisão do modelo:

$$\sum \alpha_i y_i K(x_i, x_j) + b \tag{16}$$

Onde por conta de y_i, a soma é realizada em cima dos vetores de suporte (support vectors), que são as amostrar mais próximas do divisor.
Como é um método supervisionado, devemos ainda achar os melhores valores para as constantes α_i, y_i, b e γ dado um conjunto de dados de treinamento. Para isso pode-se utilizar técnicas como mínimos quadrados e validação cruzada.
Outrossim, esse método pode ser utilizada para regressão, onde a função de regressão vai ser:

$$\sum (\alpha_i - \alpha_i *) K(x_i, x_j) + b \tag{17}$$

Algumas variáveis podem entrar nesse modelo, como o peso de error para as margens.

Finalmente, é importante citar que quando se tem mais de duas classes em um conjunto de dados, é necessário usar o método de classificação "one-vs-one"ou "one-vs-rest"para resolver o problema.
Iremos utilizar a implementação do SVM em código via scikit-learn, que é uma biblioteca de código aberto em python. Esse repositório fornece toda a implementação matemático, facilitando a prática desse método. É possível encontrar a documentação aqui: `https://scikit-learn.org/stable/modules/svm.html`