

平面支持向量机 (SVM, Support Vector Machine) 是一种常用的二分类算法，主要目的是寻找一个最优的决策边界（超平面），使得该边界能够有效地将不同类别的数据点分开，并且具有最好的泛化能力。

## 支持向量机的核心思想

支持向量机的目标是找到一个最优的超平面，使得两类数据的间隔（即类别之间的边界）最大化。该超平面能够有效地将训练数据分成两类，同时使得每类中的点尽可能远离这个超平面，从而增强模型的鲁棒性。

## 关键概念

### 1. 超平面 (Hyperplane) :

- 在二维空间中，超平面就是一条直线；在三维空间中，超平面是一个平面；在更高维的空间中，超平面是一个维度比空间维度少1的超平面。
- 设超平面的方程为： $[w \cdot x + b = 0]$  其中， $(w)$  是法向量（垂直于超平面的向量）， $(x)$  是样本点， $(b)$  是偏置项。

### 2. 支持向量 (Support Vectors) :

- 支持向量是距离超平面最近的那些训练样本。这些样本在确定决策边界时起到了决定性作用。

### 3. 间隔 (Margin) :

- 间隔是指从超平面到支持向量的距离。支持向量机的目标是最大化间隔，这样可以使得分类器对新数据的泛化能力更强。
- 假设对于类别1的数据点 ( $y = +1$ ) 和类别2的数据点 ( $y = -1$ )，超平面与类别1和类别2的支持向量之间的距离分别为1，那么最大间隔为2。

## 数学公式和推导

### 1. 最优超平面的确定： 设训练数据集为 $\{(x_i, y_i)\}_{i=1}^n$ ，其中 $(x_i \in \mathbb{R}^d)$ 为第 $(i)$ 个样本， $(y_i \in \{-1, +1\})$ 为标签。目标是找到超平面，使得每个样本点 $(x_i)$ 满足以下约束条件：

- 对于类别1的样本， $(w \cdot x_i + b \geq 1)$
- 对于类别2的样本， $(w \cdot x_i + b \leq -1)$

这些条件可以合并成： $[y_i (w \cdot x_i + b) \geq 1, \quad \text{forall } i = 1, 2, \dots, n]$  这表示每个样本点都被正确分类，并且距离超平面至少为1。

### 2. 最大化间隔： 支持向量机的目标是最大化间隔，即最小化 $(\frac{1}{|w|})$ ，其中 $(|w|)$ 是法向量的长度。

为了最大化间隔，目标是最小化下面的目标函数： $[\min_w \frac{1}{2} |w|^2]$  这是一个凸优化问题，结合约束条件  $(y_i (w \cdot x_i + b) \geq 1)$ ，可以通过拉格朗日乘数法求解最优解。

### 3. 优化问题： 通过拉格朗日对偶变换，最终得到如下的优化问题： $[\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j]$ 其中， $(\alpha_i)$ 是拉格朗日乘数，满足约束 $(\alpha_i \geq 0)$ 且 $(\sum_{i=1}^n \alpha_i y_i = 0)$ 。

## 求解过程

### 1. 确定超平面的参数： 通过优化问题可以求得 $(\alpha_i)$ 的值，进而得到超平面参数 $(w)$ 和偏置项 $(b)$ 。

2. **分类预测**: 给定一个新的样本  $(x)$ , 其预测结果为:  $[\text{sign}(w \cdot x + b)]$  其中,  $(\text{sign}(z))$  是符号函数, 若  $(z > 0)$ , 则分类为  $+1$ ; 若  $(z < 0)$ , 则分类为  $-1$ 。

例子：二维空间中的线性可分情况

假设我们有如下数据集，其中每个样本的标签为  $(\pm 1)$ :

$(x_1)$	$(x_2)$	标签 $(y)$
2	3	+1
3	3	+1
6	5	-1
7	8	-1

我们希望找到一条直线（超平面）来将这些点分开，使得它们之间的间隔最大。

1. **构造优化问题**: 目标是最大化间隔，最小化  $(\frac{1}{2} |w|^2)$ ，同时满足:  $[y_i (w \cdot x_i + b) \geq 1]$  对于所有的  $(i)$ 。
2. **求解**: 通过求解这个优化问题，可以得到最优的  $(w)$  和  $(b)$ ，从而确定最优的超平面。
3. **分类**: 对于新来的数据点  $(x = (x_1, x_2))$ ，我们根据  $(w \cdot x + b)$  的符号来进行分类。

线性不可分的情况

当数据不是线性可分时，支持向量机通过引入 **软间隔** 来处理。即允许部分数据点违反分类约束，这样的点会有一个 **松弛变量**  $(\xi_i)$ ，表示允许的误差。这时的优化问题会变成:  $[\min_w \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i]$  其中,  $(C)$  是一个正则化参数，控制松弛变量的惩罚程度。增加  $(C)$  会导致对分类错误的容忍度减小，模型会更加“严格”。

总结

支持向量机的核心思想是寻找一个最优的超平面，该超平面能最大化不同类别之间的间隔，从而提高模型的泛化能力。SVM 是一种强大的分类算法，在处理高维数据和复杂模式时表现优秀，尤其适用于线性可分或通过核方法可以转化为线性可分的情形。