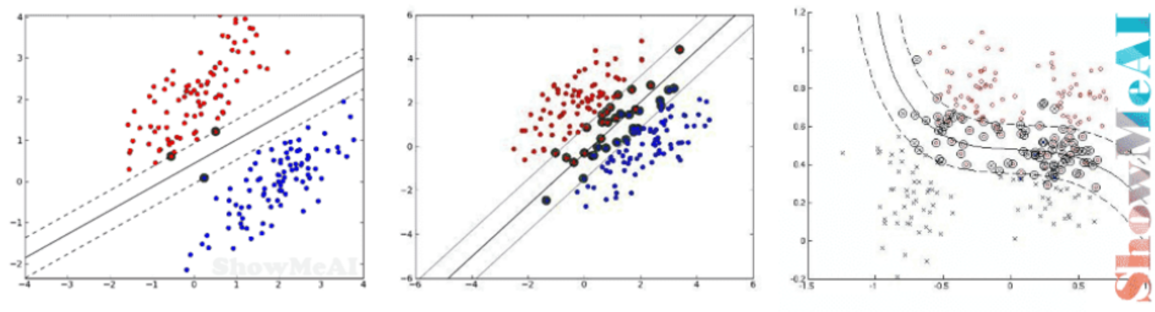
SVM

分类



线性可分支持向量机
linear support vector machine
in linearly separable case
亦称作硬间隔支持向量机

线性支持向量机
linear support vector machine
又叫软间隔支持向量机

非线性支持向量机
non-linear
support vector machine

1. 线性可分支持向量机：训练数据线性可分的情况下，通过硬间隔最大化(hard margin maximization)学习一个线性的分类器
2. 线性支持向量机：训练数据近似线性可分的情况下，通过软间隔最大化(soft margin maximization)学习一个线性的分类器
3. 非线性支持向量机：训练数据线性不可分的情况下，通过使用核技巧(kernel trick)以及软间隔最大化，学习非线性分类器

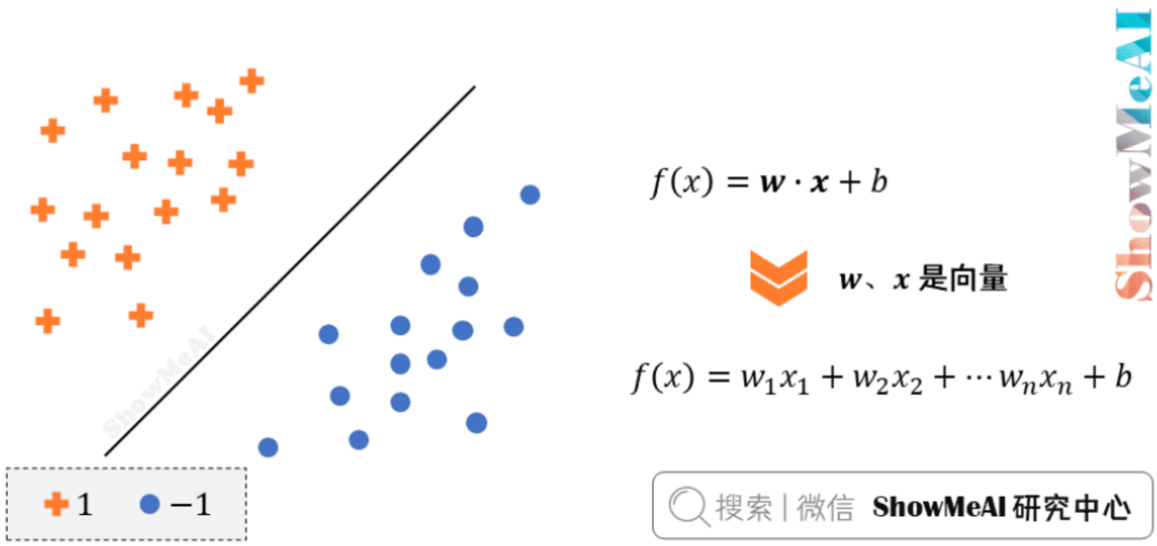
> 核函数(kernel function)：表示将输入从输入空间映射到特征空间得到的特征向量之间的内积
>
> 核技巧(kernel trick)：使用核函数学习非线性支持向量机，等价于隐式地在高维的特征空间中学习线性支持向量机
>
> SVM是一个二分类模型

最大间隔分类器

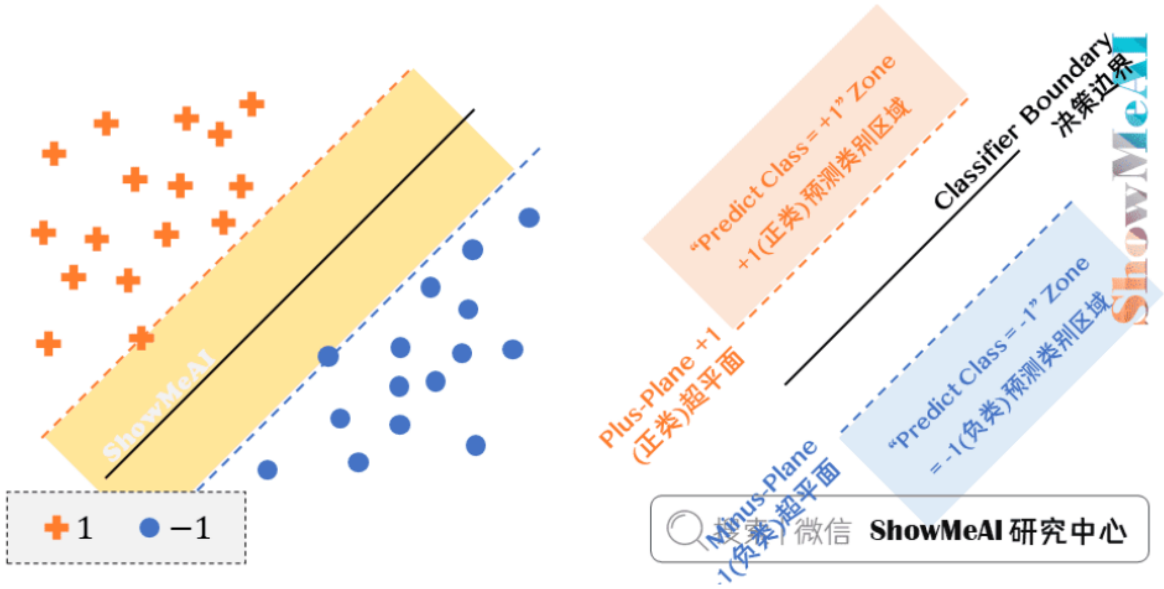
决策边界有无数个，但是要找到与两侧**最近**的数据点有**最大**距离的决策边界，这样的边界有最大的**容错性**。

SVM详解

线性可分SVM与硬间隔最大化



> 边界$f(x) = \mathbf{w} \cdot \mathbf{x} + b$，其中w与x都是向量，公式等价于
> $f(x) = w_1x_1 + \cdots + w_nx_n + b$



> 红色和蓝色的线是支持向量(support vector)所在的面，红色和蓝色线之间的间隙就是要最大化的分类间的间隙M(Margin Width)，$M = \frac{2}{||w||}$

定义：特征空间训练数据集$T = (x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$，其中$x_i \in R$，$y_i \in \{+1, -1\}$。$x_i$代表第$i$个特征向量；$y_i$是类标记，$+1$为正例，$-1$为负例

几何间隔

超平面$wx + b = 0$，超平面关于样本点(x_i, y_i)的几何间隔为$\gamma_i = y_i(\frac{w}{||w||} \cdot x_i + \frac{b}{||w||})$，超平面关于所有样本点的几何间隔的最小值为$\gamma = \min\limits_{i=1,2,\cdots,N} \gamma_i$(即为 支持向量到超平面的距离)

1. SVM模型的求解最大分割超平面问题等价为：$\max\limits_{w,b} \gamma \ s.t. \ y_i(\frac{w}{||w||} \cdot x_i + \frac{b}{||w||}) \geq \gamma$，$s.t.$代表$subject \ to$，即为约束条件
2. 将约束条件两边同时除以γ：$y_i(\frac{w}{||w||\gamma} \cdot x_i + \frac{b}{||w||\gamma}) \geq 1$
3. 令$w = \frac{w}{||w||\gamma}$，$b = \frac{b}{||w||\gamma}$：$y_i(w \cdot x_i + b) \geq 1$
4. 最大化γ等价于最大化$\frac{1}{||w||}$，等价于最小化$\frac{1}{2}||w||^2$。因此SVM求解最大分割超平面问题化成：$\min\limits_{w,b} \frac{1}{2}||w||^2 \ s.t. \ y_i(w \cdot x_i + b) \geq 1$

对偶算法

将求解线性可分支持向量机的最优化问题转换成 对偶问题(dual problem) 来求解

1. **构建拉格朗日函数(Lagrange function)**：

> $L(w,b,\alpha)=\frac{1}{2}w^Tw+\sum\limits_{i=1}^N\alpha_i(1-y_i(w^Tx_i+b))=\frac{1}{2}w^Tw-\sum\limits_{i=1}^N\alpha_iy_i(w^Tx_i+b)+\sum\limits_{i=1}^N\alpha_i$，其中$\alpha_i\geq 0$是拉格朗日乘子。根据拉格朗日对偶性，原始问题等价于极大极小问题$\max\limits_{\alpha}\min\limits_{w,b}L(w,b,\alpha)$

2. 求$L(w,b,\alpha)$对w,b的极小值：

> $\frac{\partial L}{\partial w}=0\rightarrow w=\sum\limits_{i=1}^N\alpha_iy_ix_i$，$\frac{\partial L}{b}=0\rightarrow\sum\limits_{i=1}^N\alpha_iy_i=0$
>
> 代入到$L(w,b,\alpha)$可得：
>
> $\min\limits_{w,b}L(w,b,\alpha)=\frac{1}{2}(\sum\limits_{i=1}^N\alpha_iy_ix_i)^T\sum\limits_{i=1}^N\alpha_iy_ix_i-\sum\limits_{i=1}^N\alpha_iy_i((\sum\limits_{i=1}^N\alpha_iy_ix_i)^Tx_i+b)+\sum\limits_{i=1}^N\alpha_i$
>
> 化简为$\min\limits_{w,b}L(w,b,\alpha)=-\frac{1}{2}\sum\limits_{i=1}^N\sum\limits_{j=1}^N\alpha_iy_i\alpha_jy_jx_i^Tx_j+\sum\limits_{i=1}^N\alpha_i$

3. 求$\min\limits_{w,b}L(w,b,\alpha)$对$\alpha$的极大值：

> $\max\limits_{\alpha}(-\frac{1}{2}\sum\limits_{i=1}^N\sum\limits_{j=1}^N\alpha_iy_i\alpha_jy_jx_i^Tx_j+\sum\limits_{i=1}^N\alpha_i)\ s.t.\ \sum\limits_{i=1}^N\alpha_iy_i=0$
>
> 等价为$\min\limits_{\alpha}\frac{1}{2}\sum\limits_{i=1}^N\sum\limits_{j=1}^N\alpha_iy_i\alpha_jy_jx_i^Tx_j-\sum\limits_{i=1}^N\alpha_i\ s.t.\ \sum\limits_{i=1}^N\alpha_iy_i=0$

4. 假设对偶最优化问题对α的解为$\alpha^\star$：

> 求得$w^\star=\sum\limits_{i=1}^N\alpha_i^\star y_ix_i$
>
> 根据KKT条件：$\begin{cases}\text{乘子非负}:\alpha_i\geq 0\\ \text{约束条件}:y_i(w\cdot x_i+b)\geq 1\\ \text{互补条件}:\alpha_i(y_i(x_i^Tw+b)-1)=0\end{cases}$
>
> 至少存在一个j使得$y_j(x_j^Tw^\star+b^\star)-1=0$，因此$b^\star=y_j-\sum\limits_{i=1}^N\alpha_i^\star y_ix_j^Tx_i$(注意$y_j^2=1$)

5. 分离超平面可以写成：

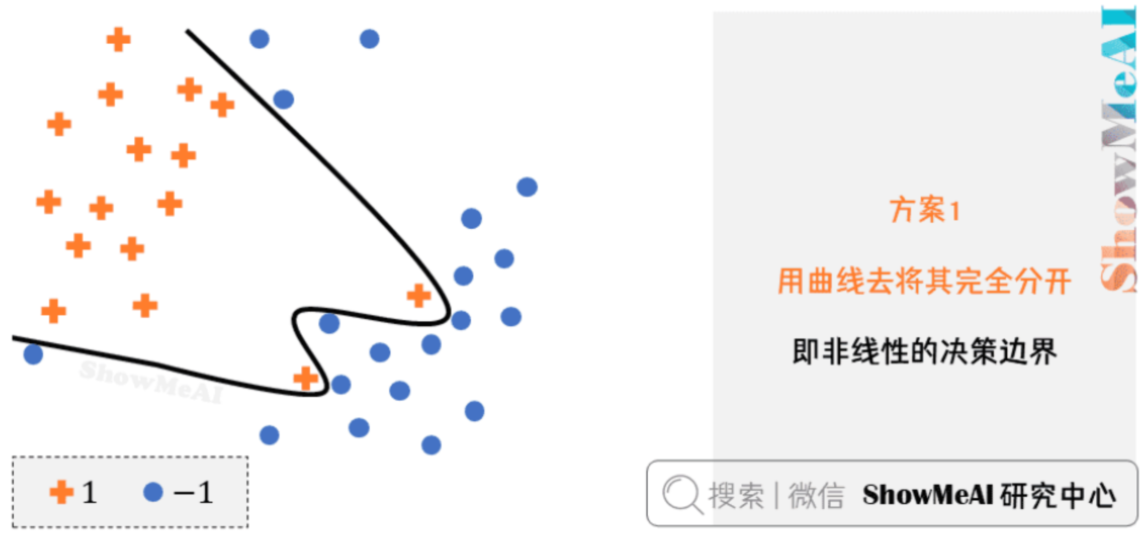
> $\sum\limits_{i=1}^N\alpha_i^\star y_i(x\cdot x_i)+b^\star=0$
>
> 决策函数只依赖于输入**x**和训练样本输入x_i的内积

6. 分类决策函数可以写成：

$$f(x)=sign(\sum_{i=1}^{N}\alpha_i^{\star}y_i(x\cdot x_i)+b^{\star})$$

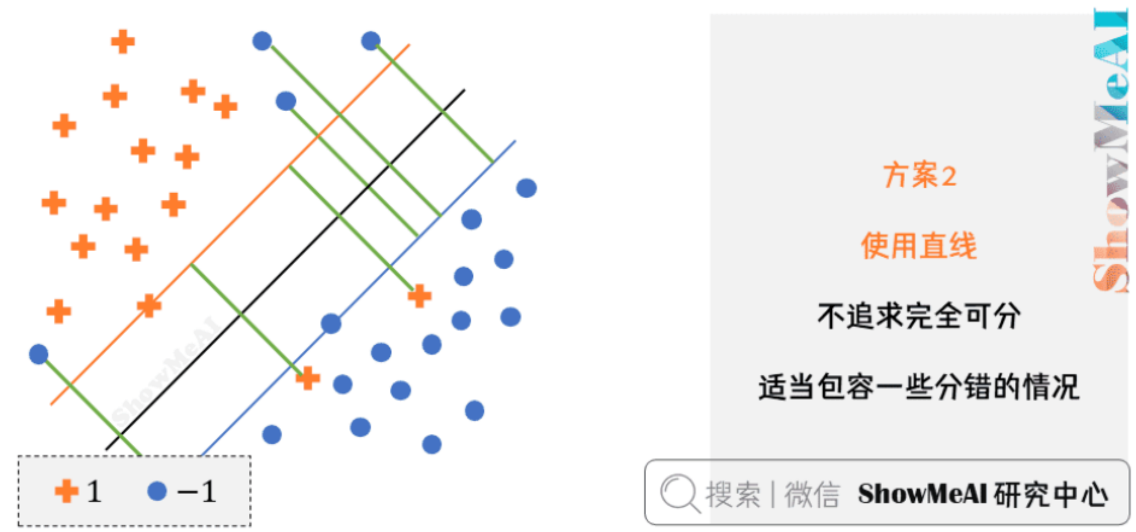
线性SVM与软间隔最大化

方案1



用曲线将其完全分开(非线性决策边界)

方案2



使用直线，不追求完全可分，但加入惩罚函数(分错点到正确未知的距离)

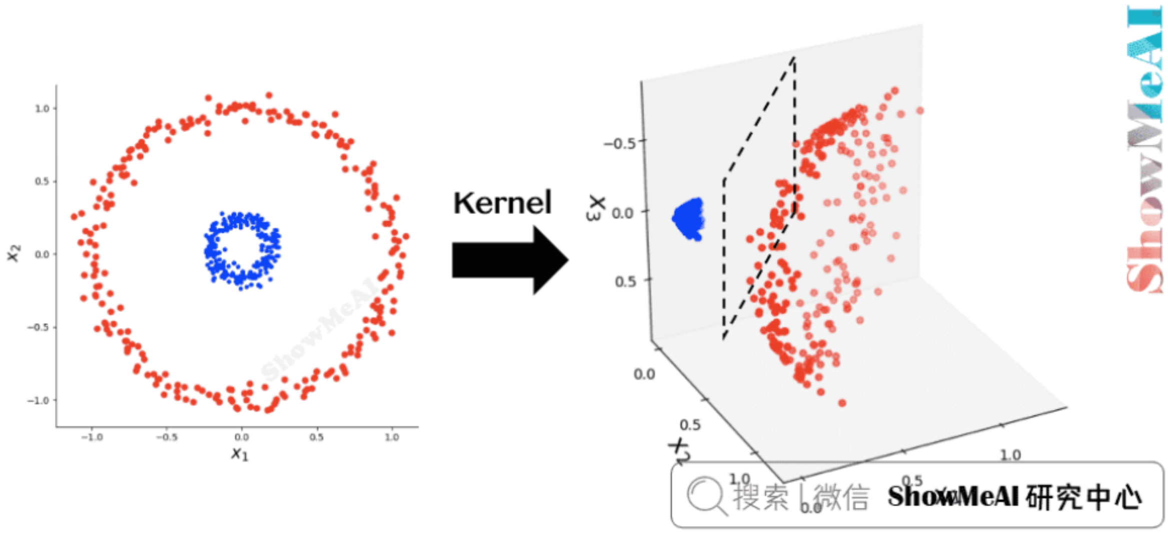
图中绿色的线代表分错点到其相应的决策面距离

则问题可化为：$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{R}\epsilon_i \ \ s.t. \ \ y_i(w^Tx_i+b) \geq 1-\epsilon_i$，其中$R$为全部点的数目，$C$是由用户指定的系数(表示对分错点加入多少的惩罚)，当x_i在正确一边时ϵ为0

同样可以求解一个拉格朗日对偶问题，得到原问题的对偶问题的表达式：
$\max\limits_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j \ \ s.t. \ \ 0 \leq \alpha_i \leq C$，$\sum_{i=1}^{N}\alpha_i y_i = 0$

非线性SVM与核函数

无法近似可分的情况下，可以将样本从原始空间映射到更高维的特征空间，使得样本在这个特征空间内线性可分，然后用SVM进行求解。



对偶问题：$\max\limits_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j \ \ s.t. \ \ 0 \leq \alpha_i \leq C$，$\sum_{i=1}^{N}\alpha_i y_i = 0$

令$x_i^T x_j = \kappa(x_i, x_j)$，这个式子作用就是将线性的空间映射到高维的空间。

常见的有多项式核$\kappa(x_i, x_j) = (x_i \cdot x_j + 1)^d$，高斯核$\kappa(x_i, x_j) = e^{-\frac{(x_i-x_j)^2}{2\sigma^2}}$