第十四周学习笔记—支持向量机理论

2022-07-15

1. 分类决策函数引入

我们得到的线性可分支持向量机如下:

定理. 线性可分支持向量机

给定线性可分训练数据集，通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为:

$$\omega^* \cdot x + b^* = 0$$

以及相应的分类决策函数:

$$f(x) = sign(\omega^* \cdot x + b^*)$$

称为线性可分支持向量机。

看着我们的分类决策函数，$\omega \cdot x + b$ 除了与我们的感知机模型比较相似，还与 Logistic 回归相似。

提出问题，支持向量机与感知机或者逻辑回归有什么关系吗？

我们回顾一下之前学习的 Logistic 回归。回归模型其实是一个广义的线性模型。如果我们想把所有的样本分为两类，一类是 0，一类是 1，我们用 y 表示，则有 $y \in \{0, 1\}$，我们有下面的等式:

$$\log \frac{p(Y=1|x)}{1-p(Y=1|x)} = \omega \cdot x + b$$

其中 $\omega \cdot x$ 表示为向量的内积。

我们的重点是要求得 ω 和 b 两个参数。

因为涉及概率，我们可以尝试极大似然法:

我们的训练数据集为:

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$$

那么根据之前的学习，我们的对数似然函数表示为:

$$L(\omega, b) = \sum_{i=1}^{N} [y_i(\omega \cdot x_i + b) - \log(1 + e^{\omega \cdot x_i + b})]$$

我们知道参数空间为 $N+1$ 维，我们可以使用遍历求解，网格搜索得到最佳参数。但是这钟方法比较麻烦，耗时较大。

我们可以计算解析解:

$$\frac{\partial L}{\partial \omega} = 0$$
$$\frac{\partial L}{\partial b} = 0$$

但是解析解表示的方法很麻烦, 不适合。

我们最后也可以使用迭代法。

综上所述，我们可以采用的计算方法有遍历求解、解析解以及迭代法。但是这里的求解我们都没有考虑几何意义。

2.Logistic 回归

在了解几何意义之前，我们先看一下 Logistic 回归是如何做分类的。

给定一个实例点 x，有下面的两种情况:

$$p(Y=1|x) \geq 0.5, x \to Y=1$$
$$p(Y=1|x) < 0.5, x \to Y=0$$

我们可以得出:

$$\begin{aligned} & p(Y=1|x) > 0.5 \Rightarrow \frac{p(Y=1|x)}{1-p(Y=1|x)} > 1 \\ & \Rightarrow \log \frac{p(Y=1|x)}{1-p(Y=1|x)} > 0 \\ & \Rightarrow \omega \cdot x + b > 0 \end{aligned}$$

于是我们得出的分类为:

$$Y = \begin{cases} 1, \omega \cdot x + b \geq 0 \\ 0, \omega \cdot x + b < 0 \end{cases}$$

3.Logistic 回归到感知机

我们在感知机中 Y 的分类为 $+1$ 和-1，即正类和负类。

我们改变逻辑回归中的分类:

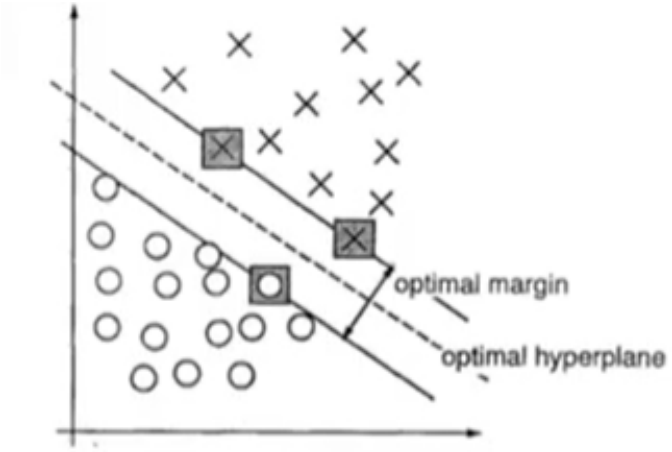
$$Y = \begin{cases} 1, \omega \cdot x + b \geq 0 \\ -1, \omega \cdot x + b < 0 \end{cases}$$

这样我们就实现了从 Logistic 回归变成了具有一定几何意义的感知机模型。

我们可以总结出感知机模型的几何意义：在特征空间中寻找一个分离超平面，这个超平面将特征空间划分成两个部分，并且 $+1$ 类和-1 类样本点尽量位于超平面的两侧。如果想要 $+1$ 类和-1 类样本严格位于超平面两侧，特征空间需要满足线性可分。

4. 支持向量机名称的由来

支持向量机的英文名称为:support vector machines，简称为 SVM。



为什么叫支持向量机？按照感知机的想法，我们只需要找到一个超平面，将上下两个平面分开即可。

在感知机中，我们有一个集合 M，其中有未分类点 $x_i \in M$，此时我们的目标函数为:

$$\min[-\sum_{x_i \in M} y_i(\omega \cdot x_i + b)]$$

但是，面对问题，我们用上述方法得到的分离超平面不唯一。

那我们选取哪些点最为合适呢？涉及到分类确信度，采用的度量因素分为距离和分类正确性。

首先来看一下距离度量因素，任一个分类实例点 (x_i, y_i) 到分离超平面的距离为:

$$\frac{|\omega \cdot x_i + b|}{||w||}$$

度量完距离因素，我们还需要度量分类是否正确，即度量分类正确性: 若分类正确，则 y_i 与 $\omega \cdot x_i + b$ 同号，否则异号。

我们把上面两个度量因素综合起来，引入几何间隔的定义:

$$\frac{y_i(\omega \cdot x_i + b)}{||\omega||}$$

我们记为 γ_i。

哪些样本点最有用呢？—距离超平面最近的点，即:

$$\min_{i=1,2,\cdots,N} \gamma_i$$

如果我们想找到分离超平面，而且是唯一的，我们自然希望把这些点分的越开越好。什么时候分的最开呢？就是把最小间隔最大化。因此，我们的目标是:

$$\max_{\omega,b} \min_{i=1,2,\cdots,N} \gamma_i$$

最有用的点，叫做支持向量，对于上图来说，支持向量是位于分离超平面边界上的点，这也是支持向量机名称的由来。

在 n 空间里，点和向量是对应的。

5. 线性可分支持向量机

我们熟悉的感知机算法中的线性可分定义为:

定理. 数据集的线性可分

给定一个数据集为:

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$$

其中:

$$x_i \in \chi = R^n, y_i \in \{+1, -1\}, i = 1, 2, \cdots, N$$

如果存在一个超平面 $\omega \cdot x + b = 0$，可以将 N 个样本点分为两类:

$$Y = \begin{cases} 1, \omega \cdot x + b > 0 \\ -1, \omega \cdot x + b < 0 \end{cases}$$

则称数据集 T 为线性可分数据集。

如果数据集线性可分，那么不存在模糊点，即没有满足 $\omega \cdot x + b = 0$ 的样本点。

线性支持向量机的想法是先找到最小的几何间隔，然后再将最小的几何间隔最大化。

一般来说，一个点距离分离超平面的远近可以表示分类预测的确信程度。我们引入几何间隔和函数间隔。

定理. 几何间隔

训练数据集 T，分离超平面 $\omega \cdot x + b = 0$，则实列点 (x_i, y_i) 到超平面之间的几何间隔为:

$$\gamma_i = \frac{|\omega \cdot x_i + b|}{||\omega||}$$

我们知道，y_i 的取值为 1 或者-1，所以当 y_i 与 $\omega \cdot x_i + b$ 同号时，我们的分类正确，我们也可以改写几何间隔为:

$$\gamma_i = \frac{y_i(\omega \cdot x_i + b)}{||\omega||}$$

其中，$||\omega|| = \sqrt{\omega_1^2 + \omega_2^2 + \cdots + \omega_N^2}$

我们这样做的目的很简单，就是为了去掉原式中的不等式，简化计算。

最后整理得到:

$$\gamma_i = y_i(\frac{\omega}{||\omega||} \cdot x_i + \frac{b}{||\omega||})$$

ω 和 b 为需要求解的参数。

我们几何间隔的最小值为:

$$\gamma = \min_i \gamma_i$$

推出我们的问题为:

$$\begin{cases} \max\limits_{\omega,b} \gamma \\ s.t. \quad y_i(\frac{\omega}{||\omega||} \cdot x_i + \frac{b}{||\omega||}) \geq \gamma \end{cases}$$

其中，约束条件中 $i = 1, 2, \cdots, N$，属于 N 个不等式问题。

用度量 $y_i(\omega \cdot x + b)$ 来表示分类的正确性及确信度，这就是函数间隔的概念。

定理. *函数间隔*

(x_i, y_i) 与超平面之间的函数间隔 $\hat{\gamma}_i = |\omega \cdot x_i + b|$, 同号时有 $\hat{\gamma}_i = y_i(\omega \cdot x_i + b)$。

函数间隔的最小值为:$\hat{\gamma} = \min_i \hat{\gamma}_i$。

我们得到的问题为:

$$\begin{cases} \max\limits_{\omega,b} \dfrac{\hat{\gamma}}{||\omega||} \\ s.t. \quad y_i(\omega \cdot x_i + b) \geq \hat{\gamma} \end{cases}$$

函数间隔表示分类预测的正确性及确信度。但是选择分离超平面时，只有函数间隔往往是不够的。因为只要成比例改变 ω 和 b，例如将它们改为 2ω 和 $2b$，超平面并没有改变，但是函数间隔却变为原来的两倍。

我们举一个例子: 对于同一个样本，不同的人得到了两个不同的分离超平面:

$$\begin{cases} 3x^{(1)} + 4x^{(2)} + 1 = 0 \\ 6x^{(1)} + 8x^{(2)} + 2 = 0 \end{cases}$$

这一例子启示我们，可以对分离超平面的法向量 ω 加某些约束，如规范化，使得 $||\omega|| = 1$，使得间隔是确定的。此时我们的分离超平面化为:

$$\frac{3}{5}x^{(1)} + \frac{4}{5}x^{(2)} + \frac{1}{5} = 0$$

这样我们得到的超平面具有可识别性。

其实，此时函数间隔成为了几何间隔。

但是我们也面临着困难，在 N 个不等式约束问题中加上一个等式约束 $||\omega|| = 1$，即 $\omega_1^2 + \omega_2^2 + \cdots + \omega_N^2 = 1$，这样会导致约束条件更加复杂。

原问题为:

$$\max_{\omega,b} \frac{\hat{\gamma}}{||\omega||}$$

如果想要维持 N 个约束条件，$||\omega|| = 1$ 不能使用，我们不妨从分子入手，如令 $\hat{\omega} = 1$。几何意义是: 如果实例点与超平面之间的函数距离大于 1，则压缩距离；如果小于 1，则拉伸距离。

我们这样得到的模型也具有可识别性。并且模型得到极大简化:

$$\begin{cases} \max\limits_{\omega,b} \dfrac{1}{||\omega||} \Leftrightarrow \min\limits_{\omega,b} ||\omega|| \\ s.t. \quad y_i(\omega \cdot x_i + b) \geq 1 \end{cases}$$

这是一个凸优化问题。

6. 凸优化问题

对于上面的问题:

$$||\omega|| = \sqrt{\omega_1^2 + \omega_2^2 + \cdots + \omega_N^2}$$

是一个凸函数。

我们定义凸优化问题:

定理. 凸优化问题

$$\begin{cases} \min\ f(\omega) \\ s.t. \quad \begin{cases} g_i(\omega) \leq 0, i = 1, 2, \cdots, k \\ h_j(\omega) = 0, j = 1, 2, \cdots, l \end{cases} \end{cases}$$

其中，$f(\omega)$ 和 $h_j(\omega)$ 是凸函数，$g_i(\omega)$ 是仿射函数。

此时，$g_i(\omega, b) = 1 - y_i(\omega \cdot x_i + b) \leq 0$，说明问题是一个凸优化问题。

我们接下来证明最优解的唯一性，将分为两个部分进行—存在性和唯一性。

首先来看存在性，这个很容易就能证明。因为我们的数据集是线性可分的，所以一定有 $||\omega|| \neq 0$，这样才能保证有解。

接下来证明唯一性，我们采用反证法:

证明. 假设我们的分离超平面存在两组解，$(\omega_1^*, b_1^*)^T, (\omega_2^*, b_2^*)^T$，即:

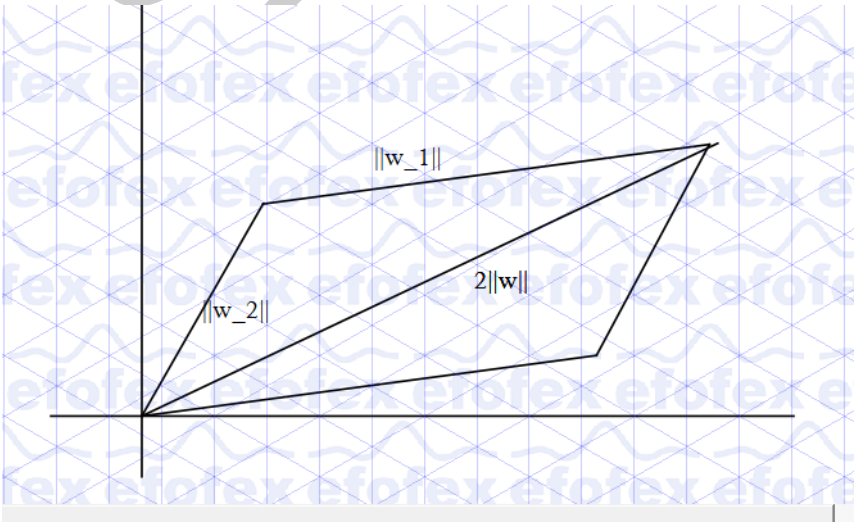
$$||\omega_1^*|| = ||\omega_2^*|| = c$$

构造:

$$\omega = \frac{\omega_1^* + \omega_2^*}{2}, b = \frac{b_1^* + b_2^*}{2}$$
$$\Rightarrow c \leq \omega = ||\frac{\omega_1^* + \omega_2^*}{2}|| = ||\frac{1}{2}\omega_1^* + \frac{1}{2}\omega_2^*|| \leq \frac{1}{2}||\omega_1^*|| + \frac{1}{2}||\omega_2^*|| = c$$

我们最终得出结论:ω_1^* 和 ω_2^* 在同一条直线上。

我们用下图辅助理解:



$$||\omega|| = \frac{1}{2}||\omega_1^*|| + \frac{1}{2}||\omega_2^*||$$
$$\Rightarrow 2||\omega|| = ||\omega_1^*|| + ||\omega_2^*||$$
$$\Rightarrow ||\omega_1^*|| + ||\omega_2^*|| \geq 2||\omega||$$

取等号时有 ω_1^* 和 ω_2^* 在同一条直线上，此时有:

$$\omega_1^* = \lambda\omega_2^*$$

我们知道 $||\omega_1^*|| = ||\omega_2^*|| = c$，所以我们知道 $\lambda = 1$ 或者 $\lambda = -1$。

若 $\lambda = -1$，则有 ω=0，$c = 0$;

若 $\lambda = 1$，则 $\omega_1^* = \omega_2^*$。

接下来我们看关于参数 b 的。

因为 $\omega_1^* = \omega_2^*$，所以我们可以得到两组解为:$(\omega^*, b_1^*), (\omega^*, b_2^*)$。

对于第一个 b_1^*，我们的分离超平面为 $\omega^* \cdot x + b_1^* = 0$，我们选取两个支持向量 $x_1^{'}, x_1^{''}$，其中前者是正类点，后者是负类点。我们代入超平面得到:

$$\begin{cases} \omega^* \cdot x_1^{'} + b_1^* = 1 \\ \omega^* \cdot x_1^{''} + b_1^* = -1 \end{cases}$$

$$\Rightarrow b_1^* = -\frac{\omega^* \cdot (x_1^{'} + x_1^{''})}{2}$$

对于第一个 b_2^*，我们的分离超平面为 $\omega^* \cdot x + b_2^* = 0$，我们选取两个支持向量 $x_2^{'}, x_2^{''}$，其中前者是正类点，后者是负类点。我们代入超平面得到:

$$\begin{cases} \omega^* \cdot x_2^{'} + b_2^* = 1 \\ \omega^* \cdot x_2^{''} + b_2^* = -1 \end{cases}$$

$$\Rightarrow b_2^* = -\frac{\omega^* \cdot (x_2^{'} + x_2^{''})}{2}$$

因为是向量内积，所以 b_1^*, b_2^* 不能直接作比。我们采用作差的方式来比较:

$$b_1^* - b_2^* = -\frac{1}{2}[\omega^* \cdot (x_1^{'} - x_2^{'}) + \omega^* \cdot (x_1^{''} - x_2^{''})]$$

我们有:

$$\begin{aligned} &\omega^* \cdot x_2^{'} + b_1^* \geq 1 = \omega^* \cdot x_1^{'} + b_1^* \\ \Rightarrow &\omega^*(x_1^{'} - x_2^{'}) \leq 0 \end{aligned}$$

$$\begin{aligned} &\omega^* \cdot x_1^{'} + b_2^* \geq 1 = \omega^* \cdot x_2^{'} + b_2^* \\ \Rightarrow &\omega^*(x_1^{'} - x_2^{'}) \geq 0 \end{aligned}$$

从上面上个不等式我们推出:

$$\omega^*(x_1^{'} - x_2^{'}) = 0$$

同理可得:

$$\omega^*(x_1^{''} - x_2^{''}) = 0$$

综上所述，我们得到:$b_1^* = b_2^*$。

所以最优解是唯一的！ □

7. 最大间隔算法

定理. 算法

输入:

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\},x_1\in R^n,y\in\{+1,-1\},i=1,2,\cdots,N$$

输出:

最大间隔分离超平面和决策函数

(1) 构建优化问题:

$$\begin{cases}\min\limits_{\omega,b}\dfrac{1}{2}||\omega||^2\\ s.t.\quad 1-y_i(\omega\cdot x_i+b)\le 0\end{cases}$$

注意: 与之前 $\min_{\omega,b}||\omega||$ 有区别，$\min_{\omega,b}||\omega||$ 等价于 $\min_{\omega,b}||\omega||^2$，$\frac{1}{2}$ 是未来求导后消掉 2。优化问题所得的解为:ω^*,b^*。

(2) 输出结果

分离超平面:

$$\omega^*\cdot x+b^*=0$$

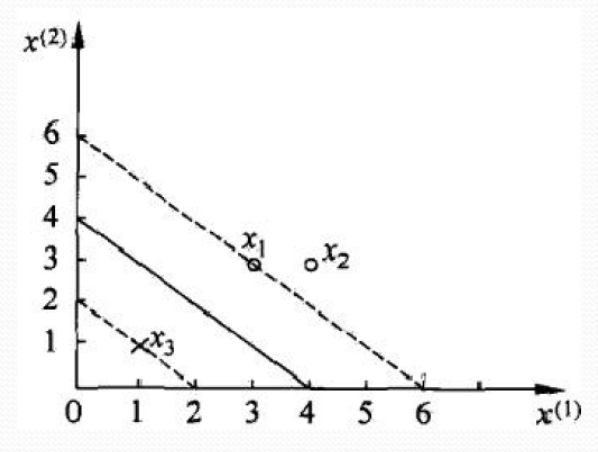
决策函数:

$$f(x)=sign(\omega^*\cdot x+b^*)$$

我们有一组实例点 (x_0,y_0)，有下面的分类情况:

$$\begin{cases}\omega^*\cdot x_0+b^*>0\Rightarrow y_0=+1\\ \omega^*\cdot x_0+b^*<0\Rightarrow y_0=-1\end{cases}$$

我们接下来看一个例子:



已知一个如上图所示的训练数据集，其正例点是 $x_1=(3,3)^T, x_2=(4,3)^T$，负例点是 $x_3=(1,1)^T$，试求最大间隔分离超平面。

证明. 解

H_1 正例所对应的超平面为:$\omega\cdot x+b=1$;H_2 负例对应的超平面为:$\omega\cdot x+b=-1$。

H_1 与 H_2 之间的距离称为最优间隔，为:$\frac{2}{||\omega||}$，H_1 与 H_2 为间隔边界。

我们设 $\omega=(\omega_1,\omega_2)^T$。

（1）优化问题

$$\min_{\omega,b}\frac{1}{2}||\omega||^2=\frac{1}{2}\omega_1^2+\frac{1}{2}\omega_2^2$$

$$s.t.\quad 1-y_i(\omega\cdot x_i+b)\le 0, i=1,2,3$$

面对这个问题，我们的约束条件分别为:

$$\begin{cases}1-1\times(3\omega_1+3\omega_2+b)\le 0\\ 1-1\times(4\omega_1+3\omega_2+b)\le 0\\ 1-(-1)\times(\omega_1+\omega_2+b)\le 0\end{cases}$$

$$\begin{cases}1-(3\omega_1+3\omega_2+b)=0\\ 1+(\omega_1+\omega_2+b)=0\end{cases}\Rightarrow \omega_1+\omega_2=1$$

我们得到:

$$\begin{aligned}\frac{1}{2}\omega_1^2+\frac{1}{2}\omega_2^2&=\frac{1}{2}(1-\omega_1)^2\\ &=\omega_1^2-\omega+\frac{1}{2}\end{aligned}$$

当 $\omega_1=\frac{1}{2}$ 时上式取得最小值，此时 $\omega_2=\frac{1}{2}$

我们得到的超平面为:

$$\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}=2$$

□

8. 线性可分支持向量机—对偶问题

优化问题:

$$\min_{\omega,b}\frac{1}{2}||\omega||^2$$

$$s.t.\quad y_i(\omega\cdot x_i+b)-1\ge 0, i=1,2,\cdots,N$$

上面的问题是包含不等式约束的凸优化问题。

我们可以找到一个广义拉格朗日函数，约束放到拉格朗日函数中，其中，α 为 N 个拉格朗日乘子。

$$\begin{aligned} L(\omega, b, \alpha) &= \frac{1}{2}||\omega||^2 + \sum_{i=1}^{N} \alpha_i (1 - y_i(\omega \cdot x_i + b)) \\ &= \frac{1}{2}||\omega||^2 - \sum_{i=1}^{N} \alpha_i y_i (\omega \cdot x_i + b) + \sum_{i=1}^{N} \alpha_i \end{aligned}$$

拆成两部分是为了方便后续对偶问题的计算。

$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N)^T$，$\alpha_i \geq 0$。

我们的原始问题为:

$$\min_{\omega,b} \max_{\alpha} L(\omega, b, \alpha)$$

对偶问题为:

$$\max_{\alpha} \min_{\omega,b} L(\omega, b, \alpha)$$

方便起见，我们可以用 $\theta_D(\alpha) = \min_{\omega,b} L(\omega, b, \alpha)$，其中 D 为对偶的含义。

对偶问题拆成两部分:

第一部分: 内部极小化 $\theta_D(\alpha)$:

$$\begin{aligned} \nabla_\omega L &= \frac{1}{2} \times 2 \cdot \omega - \sum_{i=1}^{N} \alpha_i y_i x_i \\ &= \omega - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \end{aligned}$$

$$\nabla_b L = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

从上面的梯度计算我们可以得到:

$$\begin{cases} \omega = \sum_{i=1}^{N} \alpha_i y_i x_i \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases}$$

我们代入广义拉格朗日函数为:

$$\begin{aligned} &\theta_D(\alpha) \\ \Rightarrow &\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i y_i (\sum_{j=1}^{N} \alpha_j y_j (x_j \cdot x_i)) + \sum_{i=1}^{N} \alpha_i \\ = &-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i \end{aligned}$$

第二部分: 外部极大化问题

$$\begin{cases} \max\limits_{\alpha} \ \theta_D(\alpha) \\ s.t. \begin{cases} \sum\limits_{i=1}^{N} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \\ i = 1, 2, \cdots, N \end{cases} \end{cases}$$

但是 α^* 如何得到 ω^* 和 b^*?

在解决这个问题之前，我们先看一下什么叫原始问题和对偶问题，以及求解需要满足的 KKT 条件。

在约束最优化问题中，常常利用拉格朗日对偶性将原始问题转换为对偶问题，通过解对偶问题得到原始问题的解。

定理. *原始问题*

设 $f(x), c(x), h(x)$ 是定义在 R^n 上的连续可微函数:

$$\min_{x \in R^n} \ f(x)$$

$$s.t. \begin{cases} c_i(x) \leq 0, i = 1, 2, \cdots, k \\ h_j(x) = 0, j = 1, 2, \cdots, l \end{cases}$$

引进拉格朗日函数，α_i, β_j 为乘子，$\alpha_i \geq 0$

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{i=1}^{l} \beta_j h_j(x)$$

假设给定某个 x，如果 x 违反约束条件:

$$c_i(\omega) > 0, h_j(\omega) \neq 0$$

$$\Rightarrow \theta_p(x) = \max_{\alpha, \beta: \alpha_i \geq 0} [f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)] = +\infty$$

$$\theta_p(x) = \begin{cases} f(x) \ , \ x \in s.t. \\ +\infty \ , \ otherwise \end{cases}$$

考虑极小问题:

$$\min_{x} \ \theta_p(x) = \min_{x} \ \max_{\alpha, \beta: \alpha_i \geq 0} \ L(x, \alpha, \beta)$$

与原始最优化问题等价:

$$p^* = \min_{x} \ \theta_p(x)$$

总结如下:

原始问题总结为:

$$\min_x \ \theta_p(x) = \min_x \ \max_{\alpha,\beta:\alpha_i \geq 0} \ L(x,\alpha,\beta)$$

称为广义拉格朗日函数的极小极大问题。

定义原始问题的最优值:

$$p^* = \min_x \ \theta_p(x)$$

定理. 对偶问题

定义:

$$\theta_D(\alpha,\beta) = \min_x \ L(x,\alpha,\beta)$$

则最大值问题:

$$\max_{\alpha,\beta:\alpha_i \geq 0} \theta_D(\alpha,\beta) = \max_{\alpha,\beta:\alpha_i \geq 0} \ \min_x \ L(x,\alpha,\beta)$$

上述称为广义拉格朗日函数的极大极小问题。

表示为约束最优化问题:

$$\begin{cases} \max\limits_{\alpha,\beta} \ \theta_D(\alpha,\beta) = \max\limits_{\alpha,\beta} \ \min\limits_x \ L(x,\alpha,\beta) \\ s.t. \ \alpha_i \geq 0, i = 1,2,\cdots,k \end{cases}$$

称为原始问题的对偶问题。

对偶问题的最优值:

$$d^* = \max_{\alpha,\beta:\alpha_i \geq 0} \ \theta_D(\alpha,\beta)$$

定理. 原始问题和对偶问题的关系

若原始问题和对偶问题都有最优值，则:

$$d^* = \max_{\alpha,\beta:\alpha_i \geq 0} \ \min_x \ L(x,\alpha,\beta) \leq \min_x \ \max_{\alpha,\beta:\alpha_i \geq 0} \ L(x,\alpha,\beta) = p^*$$

推论:

设 x^*，和 α^*，β^* 分别是原始问题和对偶问题的可行解，并且 $d^* = p^*$，则 x^*，和 α^*，β^* 分别是原始问题和对偶问题的最优解。

定理. KKT 条件

对原始问题和对偶问题，假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数，$h_j(x)$ 是仿射函数，并且不等式 $c_i(x)$ 是严格可行的，则 x^*，和 α^*，β^* 分别是原始问题和对偶问题的解的充

分必要条件是 x^*，和 α^*，β^* 满足 KKT 条件。

$$\nabla_x L(x^*, \alpha^*, \beta^*) = 0$$

$$\nabla_\alpha L(x^*, \alpha^*, \beta^*) = 0$$

$$\nabla_\beta L(x^*, \alpha^*, \beta^*) = 0$$

$$\alpha_i^* c_i(x^*) = 0\ ,\ i = 1, 2, \cdots, k$$

$$c_i(x^*) \leq 0\ ,\ i = 1, 2, \cdots, k$$

$$\alpha_i^* \geq 0\ ,\ i = 1, 2, \cdots, k$$

$$h_j(x^*) = 0\ ,\ j = 1, 2, \cdots, l$$

我们上式中的 $c_i(x) = 1 - y_i(\omega \cdot x_i + b)$，其中 α 相当于 ω。

$$\nabla_\omega L = \omega - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$$\Rightarrow \omega^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

b^* 应该如何表示呢？

如果 $\alpha^* = (0, 0, \cdots, 0)^T \Rightarrow \omega^* = 0$，不符合要求。

所以存在 $\alpha_j^* > 0$，使得:

$$\alpha_j^*(1 - y_j(\omega^* \cdot x_j + b^*)) = 0$$

$$\Rightarrow 1 - y_j(\omega^* \cdot x_j + b^*) = 0$$

$$\Rightarrow b^* = \frac{1}{y_j} - \omega^* \cdot x_j = \frac{1}{y_j} - \sum_{i=1}^{N} \alpha_i y_i (x_i \cdot x_j)$$

所以我们得到 b^* 的表达式为:

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

我们的超平面为:

$$\omega^* \cdot x + b^* = 0$$

我们的决策函数为:

$$f(x) = sign(\omega^* \cdot x + b^*)$$

9. 对偶算法

①: 构造优化问题

$$\min_{\alpha} \ \theta_D(\alpha) = \min_{\alpha} \ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$

$$s.t. \ \sum_{i=1}^{N}\alpha_i y_i = 0 \ , \ \alpha_i \geq 0 \ , \ i = 1, 2, \cdots, N$$

我们求得的 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^T$，N 个元素，存在 $\alpha_j^* > 0$，j 对应样本点 (x_j, y_j)。(其实我们用的 (x_j, y_j) 就是支持向量)。

②: 计算参数

$$\omega^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i$$

$$b^* = y_j - \sum_{i=1}^{N}\alpha_i^* y_i (x_i \cdot x_j)$$

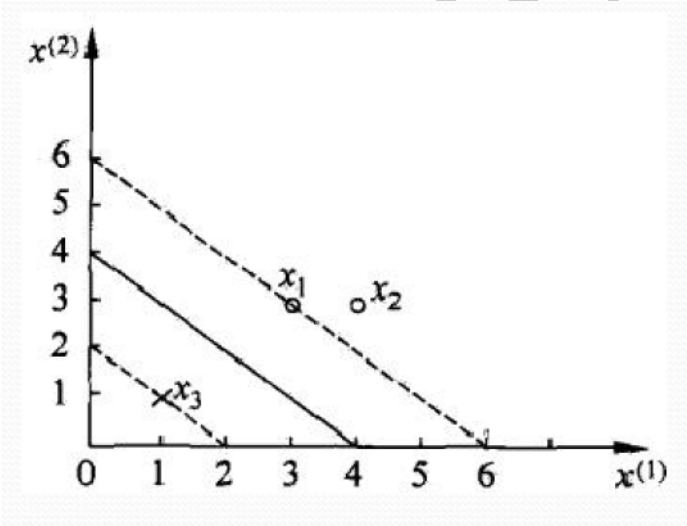
$$\Rightarrow y_j - \omega^* \cdot x_j = b^*$$

$$\Rightarrow \omega^* \cdot x_j + b^* = y_j \in \{-1, +1\}$$

我们上面等式恰好表示间隔边界。

如何找支持向量呢？

我们看一个例题:



①优化问题

$$\min\ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i$$

$$s.t.\ \sum_{i=1}^{N}y_i\alpha_i=0\ ,\ \alpha_i\ ,\ i=1,2,3$$

当 x_1 , x_1 时:

$$\alpha_1\alpha_1 y_1 y_1(x_1\cdot x_1)=18\alpha_1^2$$

当 x_1 , x_2 时:

$$2\alpha_1\alpha_2 y_1 y_2(x_1\cdot x_2)=42\alpha_1\alpha_2$$

当 x_1 , x_3 时:

$$2\alpha_1\alpha_3 y_1 y_3(x_1\cdot x_3)=-12\alpha_1\alpha_3$$

当 x_2 , x_2 时:

$$\alpha_2\alpha_2 y_2 y_2(x_2\cdot x_2)=25\alpha_2^2$$

当 x_2 , x_3 时:

$$2\alpha_2\alpha_3 y_2 y_3(x_2\cdot x_3)=-14\alpha_2\alpha_3$$

当 x_3 , x_3 时:

$$\alpha_3\alpha_3 y_3 y_3(x_3\cdot x_3)=2\alpha_3^2$$

我们代入:

$$\Rightarrow\min\ \frac{1}{2}(18\alpha_1^2+25\alpha_2^2+2\alpha_3^2+42\alpha_1\alpha_2-12\alpha_1\alpha_3-14\alpha_2\alpha_3)-(\alpha_1+\alpha_2+\alpha_3)$$

$$\sum_{i=1}^{N}y_i\alpha_i=0\Rightarrow\alpha_1+\alpha_2-\alpha_3=0$$

我们将 $\alpha_1+\alpha_2-\alpha_3=0$ 代入计算:

$$\Rightarrow\min\ 4\alpha_1^2+\frac{13}{2}\alpha_2^2+10\alpha_1\alpha_2-2\alpha_1-2\alpha_2=s(\alpha_1,\alpha_2)$$

我们需要找 $s(\alpha_1,\alpha_2)$ 的最小值。

如果求偏导数:

$$\begin{cases}\dfrac{\partial s}{\partial\alpha_1}=8\alpha_1+10\alpha_2-2=0\\[2ex]\dfrac{\partial s}{\partial\alpha_2}=13\alpha_2+10\alpha_1-2=0\end{cases}$$

我们解出:

$$\begin{cases} \alpha_1 = \dfrac{3}{2} \\ \alpha_2 = -1 \end{cases}$$

因为 α_1 , α_2 , $\alpha_3 \geq 0$，所以不满足条件。

如果不能使用费马原理，我们猜测是否在边界上。

（1）如果 $\alpha_1 = 0$

$$s(0, \alpha_2) = \frac{13}{2}\alpha_2^2 - 2\alpha_2 \Rightarrow \alpha_2 = \frac{2}{13}\ ,\ s = -\frac{2}{13}$$

（2）如果 $\alpha_2 = 0$

$$s(\alpha_1, 0) = 4\alpha_1^2 - 2\alpha_1 \Rightarrow \alpha_1 = \frac{1}{4}\ ,\ s = -\frac{1}{4}$$

我们解出 $\alpha_1 = \frac{1}{4}$, $\alpha_2 = 0$, $\alpha_3 = \frac{1}{4}$，说明 x_1 和 x_3 是支持向量。

②求解参数

$$\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i = \frac{1}{4}(3,3)^T - \frac{1}{4}(1,1)^T = (\frac{1}{2}, \frac{1}{2})^T$$

$$b^* = y_j - \omega^* \cdot x_j$$

$$\Rightarrow b^* = 1 - (\frac{1}{2}, \frac{1}{2})^T \cdot (3,3)^T = 1 - 3 = -2\ ,\ j = 1$$

$$b^* = -1 - (\frac{1}{2}, \frac{1}{2})^T \cdot (1,1)^T = -1 - 1 = -2\ ,\ j = 3$$

③

分离超平面:

$$\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2 = 0$$

决策函数为:

$$f(x) = sign(\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2)$$

10. 线性支持向量机的原始问题

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}\ ,\ x_i \in \chi \in R^n\ ,\ y_i \in \{+1, -1\}$$

线性可分:

$$y_i(\omega \cdot x_i + b) \geq 1 \quad (1)$$

$$y_i(\omega \cdot x_i + b) + \xi_i \geq 1\ ,\ \xi_i = (0,1) \quad (2)$$

$$y_i(\omega \cdot x_i + b) + \xi_i \geq 1\ ,\ \xi_i > 1 \quad (3)$$

对于 (1) 表示函数间隔 ≥ 1；对于 (2) 表示位于间隔内；对于 (3) 表示位于间隔外（异）。

上述的分类比较有弹性，所以称为软间隔:

$$y_i(\omega \cdot x_i + b) + \xi_i \geq 1$$

ξ_i 叫做弹性因子，也叫松弛变量。

目标函数:

$$\frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}\xi_i$$

上式中，c 表示松弛变量起作用的大小，又叫做惩罚系数。

当 c 较大时，对误分类惩罚大；当 c 较小时，取决于第一部分 $\frac{1}{2}||\omega||^2$，对误分类惩罚小。

此时我们的优化问题为:

$$\min_{\omega,b,\xi_i} \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}\xi_i$$

$$s.t.\ y_i(\omega \cdot x_i + b) + \xi_i \geq 1\ ,\ \xi_i \geq 0\ ,\ ,i = 1, 2, \cdots, N$$

我们接下来的目的是解 ω^* , b^*。

线性支持向量机的对偶问题:

我们的原始问题如下所示:

$$\frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}\xi_i$$

$$s.t.\ 1 - \xi_i - y_i(\omega \cdot x_i + b) \leq 0\ ,\ -\xi_i \leq 0\ ,\ i = 1, 2, \cdots, N$$

得到广义拉格朗日函数为:

$$L(\omega, b, \xi_i, \alpha, \mu) = \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i[(1 - \xi_i - y_i(\omega \cdot x_i + b))] - \sum_{i=1}^{N}\mu_i\xi_i$$

原始问题:

$$\min_{\omega,b,\xi_i}\ \max_{\alpha,\mu}\ L$$

对偶问题:

$$\max_{\alpha,\mu}\ \min_{\omega,b,\xi_i}\ L$$

（1）内部极小化 $\theta_D(\alpha,\mu)$

$$\frac{\partial L}{\partial \omega}=\omega-\sum_{i=1}^{N}\alpha_i y_i x_i=0$$

$$\frac{\partial L}{\partial b}=-\sum_{i=1}^{N}\alpha_i y_i=0$$

$$\frac{\partial L}{\partial \xi_i}=c-\alpha_i-\mu_i=0\ ,\ ,i=1,2,\cdots,N$$

$$\Rightarrow\begin{cases}\omega=\sum_{i=1}^{N}\alpha_i y_i x_i\\ \sum_{i=1}^{N}\alpha_i y_i=0\\ c-\alpha_i-\mu_i=0\ ,\ i=1,2,\cdots,N\end{cases}$$

代入计算:

$$\begin{aligned}L&=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}(c-\alpha_i-\mu_i)\xi_i+\sum_{i=1}^{N}\alpha_i\\&-\sum_{i=1}^{N}\alpha_i y_i\sum_{j=1}^{N}\alpha_j y_j(x_j\cdot x_i)-\sum_{i=1}^{N}\alpha_i y_i b\\&=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}-\sum_{i=1}^{N}\alpha_i y_i\sum_{j=1}^{N}\alpha_j y_j(x_j\cdot x_i)\\&=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i\\&=\theta_D(\alpha,\mu)\end{aligned}$$

（2）外部极大化:max $\theta_D(\alpha,\mu)$

对偶算法:

$$\min\ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}$$

$$s.t.\ \sum_{i=1}^{N}\alpha_i y_i=0\ ,\ 0\le\alpha_i\le c\ ,\ i=1,2,\cdots,N$$

其中，$\mu_i=c-\alpha_i\ge 0\Rightarrow\alpha_i\le c$

我们继续:

$$\omega^*-\sum_{i=1}^{N}\alpha_i^* y_i x_i=0$$

$$\Rightarrow\omega^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$$

对 KKT 条件再次使用:

$$\begin{cases} \alpha_i^*(1-\xi_i-y_i(\omega^*\cdot x_i+b^*))=0 \\ -\mu_i^*\xi_i^*=0 \\ 1-\xi_i^*-y_i(\omega^*\cdot x_i+b^*)\le 0 \\ \alpha_i^*\ge 0 \\ -\xi_i^*\le 0 \\ \mu_i^*\ge 0 \end{cases}$$

我们知道，存在 $\alpha_i>0$，满足软间隔的条件。发生作用的是间隔边界上的。有 $\xi_i^*=0$，$\mu_i^*>0$，$c-\mu_i^*-\alpha_i^*=0\Rightarrow\alpha_i^*<c$。

我们得到新的约束条件为:

$$s.t.\begin{cases} \sum_{i=1}^{N}\alpha_i y_i=0 \\ 0\le\alpha_i\le c \\ i=1,2,\cdots,N \end{cases}$$

我们计算 b^* 时曾用到 y_j，实际上是间隔边界上的点，满足 $\omega\cdot x+b=1$.

我们有:

$$\begin{aligned} & y_j(\omega^*\cdot x_j+b^*=1 \\ \Rightarrow & \omega^*\cdot x_j+b^*=\frac{1}{y_j}=y_j \\ \Rightarrow & b^*=y_j-\omega^*\cdot x_j=y_j-\sum_{i=1}^{N}\alpha_i^* y_i(x_i\cdot x_j) \end{aligned}$$

定理. 线性支持向量机总结

输入: 训练集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in R^n$, $y_i\in\{+1,-1\}$。

输出: 分离超平面与分类决策函数

算法:

①: 给定惩罚参数 $c>0$，构造优化问题

$$\min_{\alpha}\ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\frac{1}{2}\sum_{i=1}^{N}\alpha_i$$

$$s.t.\ \sum_{i=1}^{N}\alpha_i y_i=0\ ,\ 0\le\alpha_i\le c\ ,\ i=1,2,\cdots,N$$

②: 求解最优化问题，得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^T$$

③: 根据 α^* 进行计算

$$\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

挑出符合 $0 < \alpha_i^* < c$ 的点 (x_j, y_j) 计算:

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

④: 得到最终结果

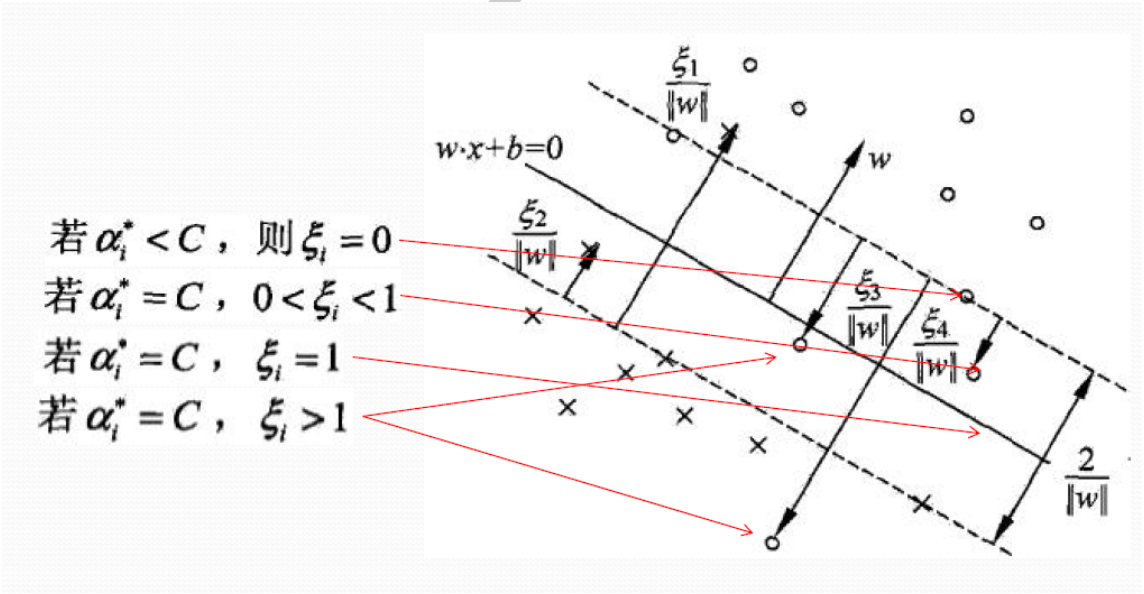
分离超平面:

$$\omega^* \cdot x + b^* = 0$$

决策函数:

$$f(x) = sign(\omega^* \cdot x + b^*)$$

我们最后提一下支持向量的定义:



11. 合页损失函数

线性支持向量机学习还有另外一种解释，就是最小化以下目标函数：

$$\sum_{i=1}^{N} [1 - y_i(\omega \cdot x_i + b)]_+ + \lambda ||\omega||^2$$

第一项:

$$L(y(\omega \cdot x + b)) = [1 - y(\omega \cdot x + b)]_+$$

称为合页损失函数。

$$[z]_+ = \begin{cases} z \ , \ z > 0 \\ 0 \ , \ z \leq 0 \end{cases}$$

线性支持向量机原始最优化问题:

$$\min_{\omega,b,\xi} \ \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}\xi_i$$

$$s.t. \begin{cases} y_i(\omega \cdot x_i + b) \geq 1 - \xi_i \ , \ i = 1, 2, \cdots, N \\ \xi_i \geq 0 \ , \ i = 1, 2, \cdots, N \end{cases}$$

等价于:

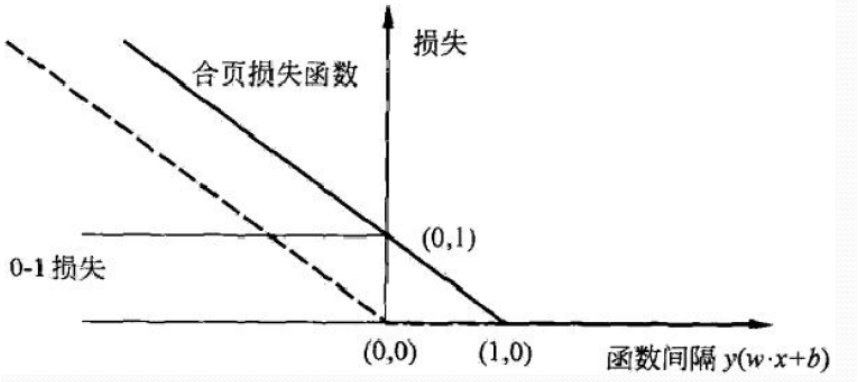
$$\min_{\omega,b} \ \sum_{i=1}^{N}[1 - y_i(\omega \cdot x_i + b)]_+ + \lambda||\omega||^2$$

我们的推导过程如下:

证明. 我们 $\xi_i = [1 - y_i(\omega \cdot x_i + b)]_+$，取合页，考虑距离大于 0 的情况。

$$\begin{aligned} &\min_{\omega,b} \ \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{N}[1 - y_i(\omega \cdot x_i + b)]_+ \\ \Rightarrow^{c>0} &\min_{\omega,b} \ \sum_{i=1}^{N}[1 - y_i(\omega \cdot x_i + b)]_+ + \frac{1}{2c}||\omega||^2 \\ \Rightarrow &\min_{\omega,b} \ \sum_{i=1}^{N}[1 - y_i(\omega \cdot x_i + b)]_+ + \lambda||\omega||^2 \end{aligned}$$

□

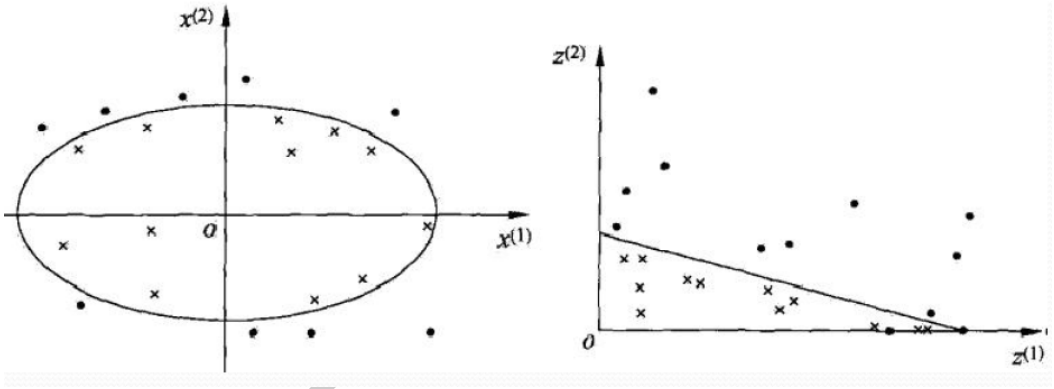


可以看到线性支持向量机损失函数作为感知机损失函数的一个上界，所以线性支持向量机可以找到那个唯一的超平面了。

12. 非线性支持向量机与核函数

线性可分: 用一个分离超平面 $\omega \cdot x + b$ 将数据集完全分开；

非线性可分: 用一个超曲面分开数据集。非线性问题往往不好求解，所以希望能用



解线性分类问题的方法解决这个问题。

采取的方法是进行一个非线性变换，将非线性问题变换为线性问题，通过解变换后的线性问题的方法求解原来的非线性问题。

即: 怎么将原空间上的点映射到新空间上的点！

13. 核函数有什么用？

原空间: 输入空间

$$\min \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$
$$s.t. \sum_{i=1}^{N}\alpha_i y_i = 0 \ , \ \alpha_i \geq 0 \ , \ i = 1, 2, \cdots, N$$

在上式中 $x_i \cdot x_j$ 是内积，新空间中可能变为 z，对应希尔伯特空间，要能计算 $z_i \cdot z_j$。

希望找到一个映射 $\phi(x) : \chi \to H$

$$z_i = \phi(x_i) \ , \ z_j = \phi(x_j)$$
$$z_i \cdot z_j = \phi(x_i) \cdot \phi(x_j) = K(x_i, x_j)$$

如果可以实现，非线性支持向量机变为:

$$\min \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i$$

接下来我们看一个例子:

证明.

$$K(x,z)=(x\cdot z)^2\ ,\ x\ z\in R^2$$

请问:$\phi\to H$?

解:

$$x=(x^{(1)},x^{(2)})^T\ ,\ z=(z^{(1)},z^{(2)})^T$$

$$\begin{aligned}K(x,z)&=(x^{(1)}z^{(1)}+x^{(2)}z^{(2)})^2\\&=(x^{(1)}z^{(1)})^2+2x^{(1)}x^{(2)}z^{(1)}z^{(2)}+(x^{(2)}z^{(2)})^2\end{aligned}$$

我们尝试 $H=R^3$

$$\phi(x)=((x^{(1)})^2,\sqrt{2}x^{(1)}x^{(2)},(x^{(2)})^2)^T$$

1)$\phi:R^2\to R^3$

$$\phi(x)\cdot\phi(z)=(x^{(1)}z^{(1)})^2+2x^{(1)}x^{(2)}z^{(1)}z^{(2)}+(x^{(2)}z^{(2)})^2=K(x,z)$$

2)

$$\phi(x)=\frac{1}{\sqrt{2}}((x^{(1)})^2-(x^{(2)})^2\ ,\ 2x^{(1)}x^{(2)}\ ,\ (x^{(1)})^2+(x^{(2)})^2)^T$$

$$\phi(x)\cdot\phi(z)=K(x,z)$$

对于同一个核函数，有多个不同的映射!

我们总结一下:

原空间:

$$\chi\in R^2\ ,\ x=(x^{(1)},x^{(2)})^T\in\chi$$

新空间:

$$Z\in R^2\ ,\ z=(z^{(1)},z^{(2)})^T\in Z$$

$$z=\phi(x)=((x^{(1)})^2,(x^{(2)})^2)^T$$

$$\Rightarrow\omega_1(x^{(1)})^2+\omega_2(x^{(2)})^2+b=0$$

$$\Rightarrow\omega_1z^{(1)}+\omega_2z^{(2)}+b=0$$

□

用线性分类方法求解非线性分类问题分为两步:

首先使用一个变换将原空间的数据映射到新空间;

然后在新空间里用线性分类学习方法从训练数据中学习分类模型。

核技巧就属于这样的方法。

核技巧应用到支持向量机，其基本想法:

通过一个非线性变换将输入空间 (欧氏空间 R”或离散集合) 对应于一个特征空间 (希尔伯特空间)，使得在输入空间中的超曲面模型对应于特征空间中的超平面模型 (支持向量机)。分类问题的学习任务通过在特征空间中求解线性支持向量机就可以完成.

14. 如何找正定核

原:$x_i \cdot x_j$

新:$\phi(x_i) \cdot \phi(x_j) = K(x_i, x_j)$ 核函数

我们要替换原来的内积定义，同一个向量它的内积一定大于等于 0，所以找正定核是最合适不过的。

证明. **解: 第一步: 我们希望找到对称函数** $K(x, z)$

$x, z \in \chi$，**为输入空间。**

需要对 $\forall x_1, x_2, \cdots, x_m \in \chi$，$K(x, z)$ **对应的** $Gram$ **矩阵半正定。**

为什么要任意选取呢？因为训练数据集不确定。

我们的 $Gram$ **矩阵为:**

原来的:

$$\begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \cdots & x_1 \cdot x_m \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \cdots & x_2 \cdot x_m \\ \cdots & \cdots & \cdots & \cdots \\ x_m \cdot x_1 & x_m \cdot x_2 & \cdots & x_m \cdot x_m \end{bmatrix}$$

新的:

$$\begin{bmatrix} K(x_1 \cdot x_1) & K(x_1 \cdot x_2) & \cdots & K(x_1 \cdot x_m) \\ K(x_2 \cdot x_1) & K(x_2 \cdot x_2) & \cdots & K(x_2 \cdot x_m) \\ \cdots & \cdots & \cdots & \cdots \\ K(x_m \cdot x_1) & K(x_m \cdot x_2) & \cdots & K(x_m \cdot x_m) \end{bmatrix}$$

我们的新矩阵需要满足半正定。

半正定的定义: 关于矩阵 A，**对** $\forall x$（非零）**存在，** $x^T A x \geq 0$，**则称矩阵** A **是半正定。**

半负定的定义:$x^T A x \leq 0$

正定的定义:$x^T A x > 0$

负定的定义:$x^TAx < 0$

我们的判定方法如下:

第一种方法:

$$x^TAx = y^TDy \geq 0$$

其中 D **为对角矩阵，所有元素均大于等于 0。**

找 A **的特征根，全部都是大于等于 0**

第二种方法:

所有主子行列式大于等于 0 □

15. 映射下的新空间

我们回顾之前学习的欧式空间的定义:

我们最初的空间叫做向量空间或者线性空间，这个空间的特点是加法运算（+）和数乘运算（×）是封闭的。

我们在此基础上定义内积定义，使得加法运算（+）、数乘运算（×）和内积运算（·）是封闭的。形成内积空间。

如果我们想知道向量的长度，我们引入范数的定义，形成赋范线性空间。

上面的向量空间、内积空间、赋范线性空间构成我们常见的欧式空间。

进一步，想要研究收敛性和极限，所有点都在空间内，叫做 $Banach$ 空间。

如果不是在我们熟知的欧氏空间中研究上面的内容，我们换了一个空间，定义了新的内积及范数，并且空间是完备的，称为希尔伯特空间。

具体的实现步骤:

证明. **解: 1):**

构成向量空间（找到向量空间）

ϕ **表示为:**$x \to K(\cdot, x)$

对于 $\forall x_i \in \chi$, $\alpha_i \in R$, $i = 1, 2, \cdots, m$

定义:

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$$

f **构成集合** S，S **变成一个向量空间。**

验证：从 S 中 $\forall f,g$，有：

$$f=\sum_{i=1}^{m}\alpha_i K(\cdot,x_i)$$
$$g=\sum_{j=1}^{v}\beta_j K(\cdot,z_j)$$

$$\begin{aligned}f+g&=\sum_{i=1}^{m}\alpha_i K(\cdot,x_i)+\sum_{j=1}^{v}\beta_j K(\cdot,z_j)\\&=\sum_{i=1}^{m+l}a_i K(\cdot,\mu_i)\in S\\af&=\sum_{i=1}^{m}a\alpha_i K(\cdot,x_i)\ ,\ a\alpha_i\in R\end{aligned}$$

综上所述，S 对加法运算和数乘运算是封闭的。

2）在集合 S 上定义内积（内积空间）

定义 $*$，对 $\forall f,g\in S$，我们定义：

$$f*g=\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_i\beta_j K(x_i,z_j)$$

内积必须满足四个条件：

$$\begin{cases}(1):(cf)*g=c(f*g)\ ,\ c\in R\\(2):(f+g)*h=f*h+g*h\ ,\ h\in S\\(3):f*g=g*f\\(4):f*f\geq 0\qquad ;f*f=0\Rightarrow f=0\end{cases}$$

我们接下来验证是否满足四个条件：

首先是第一个：

$$left\ :cf=c\sum_{i=1}^{m}\alpha_i K(\cdot,x_i)=\sum_{i=1}^{m}c\alpha_i K(\cdot,x_i)$$
$$\begin{aligned}(cf)*g&=\sum_{i=1}^{m}\sum_{j=1}^{l}(c\alpha_i)\beta_j K(x_i,z_j)\\&=c\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_i\beta_j K(x_i,z_j)\\&=c(f*g)=right\end{aligned}$$

接着是第二个：

首先我们定义:

$$h = \sum_{q=1}^{t} b_q K(\cdot, v_q)$$

令 a_i , $i = 1, 2, \cdots, m+l$ 代替 $\alpha_1, \alpha_2, \cdots, \alpha_m, \beta_1, \beta_2, \cdots, \beta_l$。

令 μ_i , $i = 1, 2, \cdots, m+l$ 代替 $x_1, x_2, \cdots, x_m, z_1, z_2, \cdots, z_l$。

$$left \ : \ (f+g) * h = \sum_{i=1}^{m+l} \sum_{q=1}^{t} a_i b_q K(\mu_i, v_q)$$

$$right \ : \ \sum_{i=1}^{m} \sum_{q=1}^{t} \alpha_i b_q K(\alpha_i, v_q) + \sum_{j=1}^{l} \sum_{q=1}^{t} \beta_j b_q K(z_j, v_q)$$

$$\Rightarrow left = right$$

我们看第三个: 很明显已经成立，因为 α_i 和 β_j 可以交换，$K(x_i, z_j)$ 是对称的，也可以换位置。

最后来看第四个:

我们首先来看第一个部分:$f * f \geq 0$

$$f * f = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j K(x_i, x_j)$$

因为 $K(x_i, z_j)$ 是对称的，所以 $Gram$ 矩阵是半正定的，有 $x^T A x \geq 0$。

所以有 $f * f \geq 0$

接下来我们看第二部分:

首先看充分性，即如果有 $f = 0$。

$$f = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i) \rightarrow \alpha_i = 0$$

$$\Rightarrow f * f = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j K(x_i, x_j) = 0$$

充分性得证!

在证明必要性之前，我们先证明接下来一个:

问题:

$$\forall f \ , \ g \in S \ , \ (f * g)^2 \leq (f * f)(g * g)$$

解: 取 $\lambda \in R$

$$f + \lambda g \in S \Rightarrow (f + \lambda g) * (f + \lambda g) \geq 0$$

$$\Rightarrow f * f + 2\lambda (f * g) + \lambda^2 (g * g) \geq 0$$

我们不妨换一下:

$$(g*g)\lambda^2+2(f*g)\lambda+f*f\geq 0$$

我们上式是不是很贴近二次函数的问题，要想恒大于 0，只需要让判别式 Δ 的值小于等于 0 即可。

即我们得知:

$$4(f*g)^2-4(g*g)(f*f)\leq 0$$
$$\Rightarrow (f*g)^2\leq (f*f)(g*g)$$

至此，我们的结论得证。接下来，我们回到原问题。

$$f=\sum_{i=1}^{m}\alpha_i K(\cdot,x_i)$$

因为 f , g 是任意取，不妨取特别的 g，令 $g=K(\cdot,x)$

$$f*g=\sum_{i=1}^{m}\alpha_i K(x,x_i)$$
$$(f*g)^2=\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j K(x,x_i)K(x,x_j)$$
$$f*f=\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j K(x_i,x_j)$$
$$g*g=K(x,x)$$
$$(f*f)(g*g)=\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j K(x_i,x_j)K(x,x)$$
$$(f*g)^2\leq (f*f)(g*g)=0$$

因为平方 $(f*g)^2\geq 0$，所以我们可以得到:

$$(f*g)^2=0$$
$$\Rightarrow f*g=0\Rightarrow\sum_{i=1}^{m}\alpha_i K(x,x_i)=0\Rightarrow\alpha_i=0\Rightarrow f=0$$

综上所述，$*$ 代表内积运算。此时我们的向量空间变为内积空间。

我们定义 f 和 g 的内积为:

$$f\cdot g=\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_i\beta_j K(x_i,z_j)$$

3) 在集合 S 上定义范数，升级希尔伯特空间

定义:

$$||f|| = \sqrt{f \cdot f}$$

此时空间转换为赋范线性空间。

新空间中 K 的特点: 再生性！如:

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$$

$$K(\cdot, x) \cdot f = \sum_{i=1}^{m} \alpha_i K(x, x_i) = f(x)$$

再比如:$K(\cdot, x) \cdot K(\cdot, z) = K(x, z)$

总结: 我们此处做的是如何从原始空间中找到一个希尔伯特空间！ □

16. 正定核函数的充要条件

设 $K : \chi \times \chi \to R$ 是对称函数，则 $K(x, z)$ 为正定核的充要条件是 $\forall x_i \in \chi$, $i = 1, 2, \cdots, m$，$K(x, z)$ 对应的 $Gram$ 矩阵 $K = [K(x_i, x_j)]_{m \times m}$ 是半正定矩阵。

接下来我们证明一下这个结论:

证明. 首先看一下充分性:

K 是半正定矩阵，我们的映射为:

$$\phi : x \to K(\cdot, x)$$

$$\chi \to H$$

$K(\cdot, x) \cdot K(\cdot, z) = K(x, z)$ 说明具有再生性，所以此时 $K(x, z)$ 为正定核。

必要性:

$K(x, z)$ 为正定核, 所以存在下面的映射:

$$\chi \to H$$

$$x \to \phi(x) \ , \ z \to \phi(z)$$

我们有:$K(\cdot, x) \cdot K(\cdot, z) = K(x, z)$

怎么判断矩阵为半正定矩阵呢?

$$\forall x_1, x_2, \cdots, x_m \in \chi$$

$$\forall c_1, c_2, \cdots, c_m \in R$$

$$c = (c_1, c_2, \cdots, c_m)^T$$

$$K = \begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_m) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_m) \\ \cdots & \cdots & \cdots & \cdots \\ K(x_m,x_1) & K(x_m,x_2) & \cdots & K(x_m,x_m) \end{bmatrix}$$

如果有:

$$c^T K c \geq 0$$

则表示 K 是半正定的矩阵。此时有:

$$c^T K c = (c_1, c_2, \cdots, c_m) \begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_m) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_m) \\ \cdots & \cdots & \cdots & \cdots \\ K(x_m,x_1) & K(x_m,x_2) & \cdots & K(x_m,x_m) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_m \end{bmatrix}$$

方法一: 我们通过内积运算

$$(c_1, c_2, \cdots, c_m) \begin{bmatrix} \phi(x_1)\cdot\phi(x_1) & \phi(x_1)\cdot\phi(x_2) & \cdots & \phi(x_1)\cdot\phi(x_m) \\ \phi(x_2)\cdot\phi(x_1) & \phi(x_2)\cdot\phi(x_2) & \cdots & \phi(x_2)\cdot\phi(x_m) \\ \cdots & \cdots & \cdots & \cdots \\ \phi(x_m)\cdot\phi(x_1) & \phi(x_m)\cdot\phi(x_2) & \cdots & \phi(x_m)\cdot\phi(x_m) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_m \end{bmatrix}$$

$$= (c_1, c_2, \cdots, c_m) \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \cdots \\ \phi(x_m) \end{bmatrix} [\phi(x_1), \phi(x_2), \cdots, \phi(x_m)] \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_m \end{bmatrix} = || \sum_{i=1}^{m} c_i \phi(x_i) ||^2 \geq 0$$

方法二: 直接表示二次型

$$\begin{aligned} \sum_{i,j=1}^{m} c_i c_j K(x_i, x_j) &= \sum_{i,j=1}^{m} c_i c_j \phi(x_i) \cdot \phi(x_j) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} [c_i \phi(x_i)] \cdot [c_j \phi(x_j)] \\ &= [c_1\phi(x_1) + c_2\phi(x_2) + \cdots + c_m\phi(x_m)] \cdot [c_1\phi(x_1) + c_2\phi(x_2) + \cdots + c_m\phi(x_m)] \\ &= || \sum_{i=1}^{m} c_i \phi(x_i) ||_2^2 \end{aligned}$$

正定核的等价定义:

设 $\chi \in R^n$, $K(x,z)$ 是定义在 $\chi \times \chi$ 上的对称函数，如果对 $\forall x_i \in \chi$, $i = 1, 2, \cdots, m$，$K(x,z)$ 对应的 $Gram$ 矩阵 $K = [K(x_i, x_j)]_{m \times m}$ 是半正定矩阵，则称 $K(x,z)$ 是正定核。 □

17. 常用核函数

①: 定义在欧式空间上

1）多项式核函数

$$K(x,z) = (x \cdot z + 1)^p$$

它的一般形式为:

$$(x \cdot z + c)^p$$

其中 c 是常量。

决策函数:

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

2）高斯核函数

$$K(x,z) = exp(-\frac{||x-z||^2}{2\sigma^2})$$

决策函数:

$$f(x) = sign(\sum_{i=1}^{N} a_i^* y_i exp(-\frac{||x-x_i||^2}{2\sigma^2}) + b^*)$$

②: 定义在离散数据集合上

字符串对应的空间映射到高维空间:

$$[\phi_n(s)]_\mu = \sum_{i, s(i)=\mu} \lambda^{(i)}$$

n 表示字符串长度，s 是字符串，$l(i)$ 表示小的字符串对应的长度。

例子: 一个文本 ['big','pig','bag']

长度为 2 的子字符串为 ['bi','bg','ig','pi','pg','ba','ag']

投影的特征空间取 R^7，计算长度:（最后一个元素位置）-（最前一个位置）+1

	bi	bg	ig	pi	pg	ba	ag
big	λ^2	λ^3	λ^2	0	0	0	0
pig	0	0	λ^2	λ^2	λ^3	0	0
bag	0	λ^3	0	0	0	λ^2	λ^2

我们得出结果:

$$K(big\ ,\ pig) = \lambda^4\ ,\ K(big\ ,\ bag) = \lambda^6$$

度量两个字符串之间的相似度的方法可以使用—余弦相似度:

$$\cos\theta = \frac{x \cdot y}{||x||||y||}$$

比较两个文本之间的相似度，如:

$$\begin{aligned} &\frac{K(big, pig)}{||K(big, big)||||K(pig, pig)||} \\ =&\frac{\lambda^4}{\sqrt{\lambda^6 + 2\lambda^4}\sqrt{\lambda^6 + 2\lambda^4}} \\ =&\frac{\lambda^4}{\lambda^6 + 2\lambda^4} = \frac{1}{2 + \lambda^2} \end{aligned}$$

2）字符串核函数

$$[\phi_n(s)]_\mu = \sum_{i, s(i)=\mu} \lambda^{(i)}$$

18. 总结

线性支持向量机

定理. 线性支持向量机

输入: 训练集 $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$，其中，$x_i \in \chi \in R^n$ ，$y_i \in \{-1, +1\}$。

输出: 分离超平面与分类决策函数

算法:

①给定惩罚系数 $c \geq 0$，构造优化问题

$$\min_\alpha \ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$

$$s.t.\ \sum_{i=1}^{N}\alpha_i y_i = 0\ ,\ 0 \leq \alpha_i \leq c\ ,\ i = 1, 2, \cdots, N$$

②求解最优化问题，得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^T$$

③根据 α^* 求解

$$\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

挑出符合 $0 < \alpha_i^* < c$ 的点 (x_j, y_j) 计算:

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

④

分离超平面为:

$$\omega^* \cdot x + b^* = 0$$

决策函数为:

$$f(x) = sign(\omega^* \cdot x + b^*)$$

非线性支持向量机

定理. 非线性支持向量机

算法:

①给定惩罚系数 $c \geq 0$，构造优化问题

$$\min_{\alpha} \ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i$$
$$s.t. \ \sum_{i=1}^{N} \alpha_i y_i = 0 \ , \ 0 \leq \alpha_i \leq c \ , \ i = 1, 2, \cdots, N$$

②求解最优化问题，得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^T$$

③根据 α^* 求解

$$\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i K(\cdot, x_i)$$

挑出符合 $0 < \alpha_i^* < c$ 的点 (x_j, y_j) 计算:

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x_j)$$

④决策函数:

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

其中 x 为新的实例。