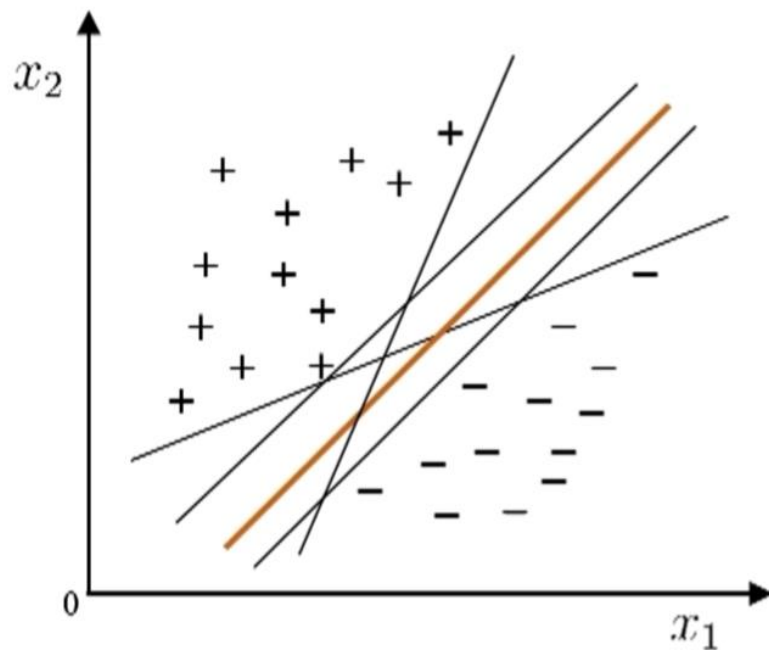




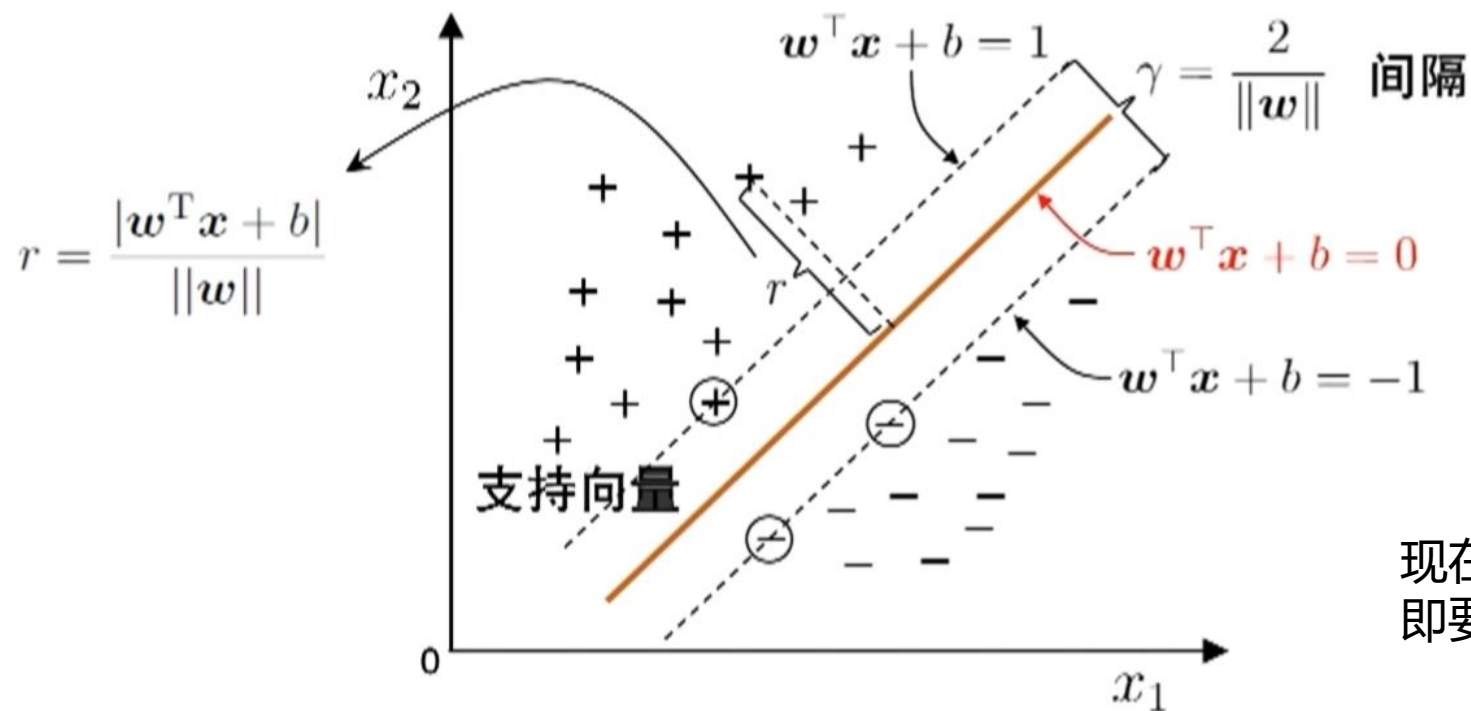
第六章 支持向量机

汇报人：何思成

将训练样本分开的超平面可能有很多, 哪一个更好呢?



考虑到噪声和干扰, 我们应该选择鲁棒性
最好的中间黄色突出的超平面进行分类



假设该超平面能够将样本正确分类，则有：

$$\begin{cases} w^T x_i + b \geq +1, & y_i = +1; \\ w^T x_i + b \leq -1, & y_i = -1. \end{cases}$$

现在要找到距离样本有最大间隔的划分超平面，即要使得 γ 最大，等价于最小化 $\|w\|^2$

超平面方程: $w^T x + b = 0$

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, m. \end{aligned}$$

- 求使得公式最小值的 w , b , 并满足下述约束条件

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

$$\text{s.t. } y_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, m.$$

- 问题相当于多元函数求极值的问题, 采用拉格朗日乘子法

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i (w^T x_i + b) - 1)$$

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

令 $L(w, b, \alpha)$ 对 w 和 b 的偏导为零可得

$$w = \sum_{i=1}^m \alpha_i y_i x_i,$$

$$0 = \sum_{i=1}^m \alpha_i y_i.$$

将 L 关于 w , b 的偏导带入式子中得到:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i (w^T x_i + b) - 1)$$

$$= \frac{1}{2} w^T w - w^T \sum_{i=1}^N \alpha_i y_i x_i - b \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i$$

$$= \frac{1}{2} w^T \sum_{i=1}^N \alpha_i y_i x_i - w^T \sum_{i=1}^N \alpha_i y_i x_i - b \cdot 0 + \sum_{i=1}^N \alpha_i$$

$$= \sum_{i=1}^N \alpha_i - \frac{1}{2} \left(\sum_{i=1}^N \alpha_i y_i x_i \right)^T \sum_{i=1}^N \alpha_i y_i x_i$$

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

$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0,$$

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, m.$$

求最大值转换成求下述约束下的最小值, α 本质上是每个训练样本的权重, 对应着训练样本 $(\mathbf{x}, y)$

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned}$$

SMO算法求解: 任意选取一对初始变量 α_i 和 α_j , 其他的 α 保持不变, 只更新求解 α_i 和 α_j , 多次迭代再求出所有的 α , 再带入下述式子求出 w 和 b 就得到了超平面方程

$$\mathbf{w}^* = \sum_{i=1}^N \alpha_i^* y_i \mathbf{x}_i$$

$$b^* = y_i - \sum_{j=1}^N \alpha_j^* y_j \mathbf{x}_j^T \mathbf{x}_i$$

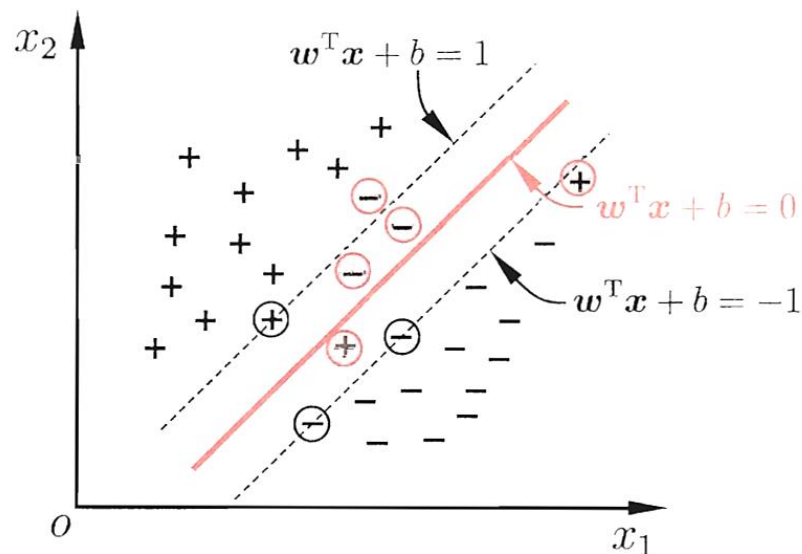
对偶问题求解出的 α 需满足KKT条件:

$$\begin{cases} \alpha_i \geq 0; \\ y_i f(\mathbf{x}_i) - 1 \geq 0; \\ \alpha_i (y_i f(\mathbf{x}_i) - 1) = 0 \end{cases}$$

于是, 对任意训练样本 $(\mathbf{x}_i, y_i)$, 总有 $\alpha_i = 0$ 或 $y_i f(\mathbf{x}_i) = 1$. 若 $\alpha_i = 0$, 则该样本将不会在式(6.12) 的求和中出现, 也就不会对 $f(\mathbf{x})$ 有任何影响; 若 $\alpha_i > 0$, 则必有 $y_i f(\mathbf{x}_i) = 1$, 所对应的样本点位于最大间隔边界上, 是一个支持向量. 这显示出支持向量机的一个重要性质: 训练完成后, 大部分的训练样本都不需保留, 最终模型仅与支持向量有关.

解出 α 后, 求出 $\mathbf{w}$ 与 b 即可得到模型

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} + b \\ &= \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \end{aligned}$$



适当放弃一些样本，也就是允许一些样本不满足约束： $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \ell_{0/1}(y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1),$$

其中 $C > 0$ 是一个常数, $\ell_{0/1}$ 是“0/1损失函数”

$$\ell_{0/1}(z) = \begin{cases} 1, & \text{if } z < 0; \\ 0, & \text{otherwise.} \end{cases}$$

采用hinge损失函数，并引入松弛因子 $\xi_i \geq 0$ ，则有：

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i$$

$$\text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, m.$$

同样，再次采用拉格朗日乘子法，则有：

$$L(\mathbf{w}, b, \alpha, \xi, \mu) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i + \sum_{i=1}^m \alpha_i (1 - \xi_i - y_i (\mathbf{w}^T \mathbf{x}_i + b)) - \sum_{i=1}^m \mu_i \xi_i$$

求L对w, b, ξ的偏导并使其为0

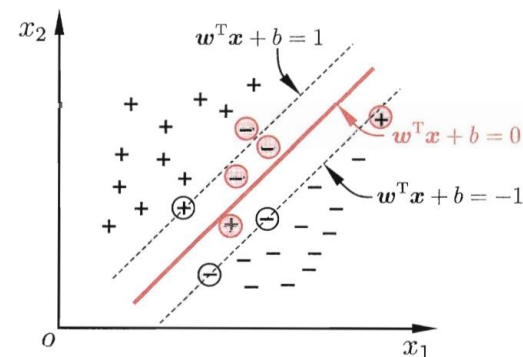
$$\begin{aligned} \Delta_{\mathbf{w}} L(\mathbf{w}, b, \xi, \alpha, \mu) &= \mathbf{w} - \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i & \mathbf{w} &= \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \\ \Delta_b L(\mathbf{w}, b, \xi, \alpha, \mu) &= - \sum_{i=1}^N \alpha_i y_i & 0 &= \sum_{i=1}^m \alpha_i y_i, \\ \Delta_{\xi_i} L(\mathbf{w}, b, \xi, \alpha, \mu) &= C - \alpha_i - \mu_i & C &= \alpha_i + \mu_i. \end{aligned}$$

令L对w, b, ξ的偏导为0后代入上式得到软间隔下的对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, m. \end{aligned}$$

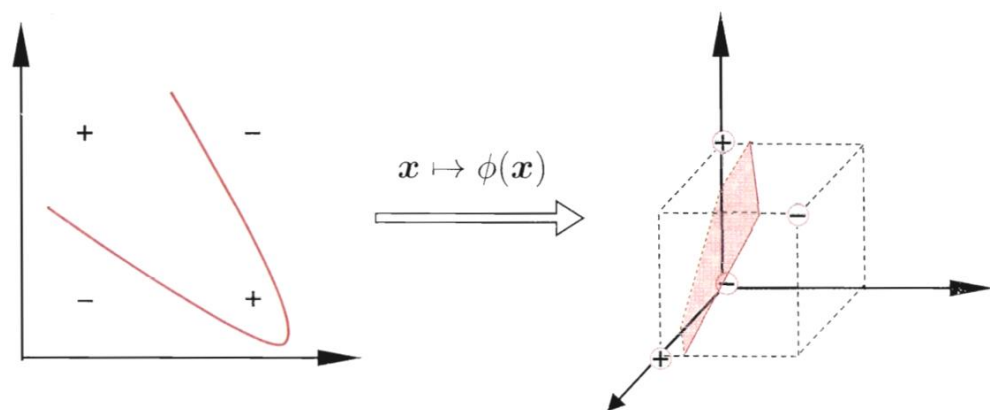
此时，软间隔支持向量机的KKT条件为：

$$\begin{cases} \alpha_i \geq 0, \quad \mu_i \geq 0, \\ y_i f(\mathbf{x}_i) - 1 + \xi_i \geq 0, \\ \alpha_i (y_i f(\mathbf{x}_i) - 1 + \xi_i) = 0, \\ \xi_i \geq 0, \quad \mu_i \xi_i = 0. \end{cases}$$



如果是非线性可分的异或问题，单个超平面没法划分

采用拉格朗日乘子法得到对偶问题：



将样本投射到高维空间，使其线性可分，模型表示为：

$$f(x) = w^T \phi(x) + b$$

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^m \alpha_i y_i = 0 ,$$

$$\alpha_i \geq 0 , \quad i = 1, 2, \dots, m .$$

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

上式中涉及高维特征向量的内积，引入核函数，设 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 在特征空间的内积等于它们在原始样本空间中通过 $\kappa(-, -)$ 计算的结果，即：

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

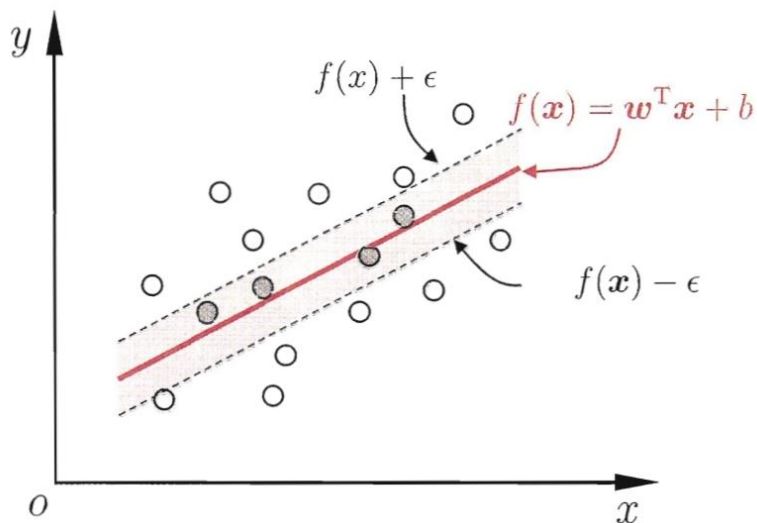
重新改写上面的式子，不必计算高维向量的内积

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & \alpha_i \geq 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

然后再来求解，则得到模型的支持向量展式：

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^T \phi(\mathbf{x}) + b \\ &= \sum_{i=1}^m \alpha_i y_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b \\ &= \sum_{i=1}^m \alpha_i y_i \kappa(\mathbf{x}, \mathbf{x}_i) + b. \end{aligned}$$

常规回归模型是计算模型输出 $f(x)$ 和真实标记 y 之间的偏差来计算损失



SVR指定 $f(x)$ 和 y 之间的差别的绝对值大于 ϵ 才计算损失，这就意味着与超平面方程的 2ϵ 间隔内则认为预测正确

针对SVR，添加损失函数并引入松弛因子，则有：

$$\min_{w, b, \xi_i, \hat{\xi}_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i)$$

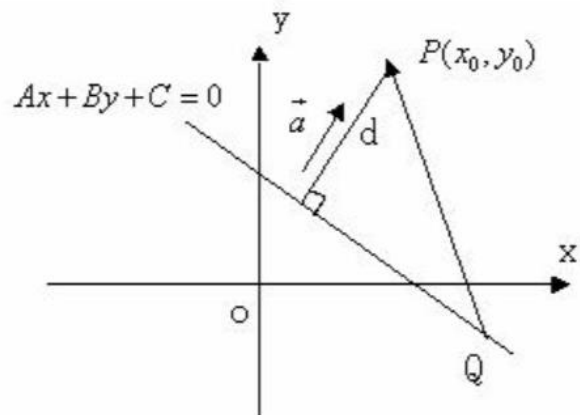
这时依然是二次规划问题，再次使用拉格朗日乘子法并通过SMO方法，递归计算 α 、 w 、 b ，得到SVR的模型方程式：

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \kappa(x, x_i) + b,$$

其中 $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 为核函数.

6.1 试证明样本空间中任意点 x 到超平面 (w, b) 的距离为式(6.2).

超平面的法向量为 ω^T , 任取平面上一点 x_0 , 则超平面可以表示为 $\omega^T \cdot x_0 + b = 0$, 一个点 x 到超平面的距离可以用该点到 x_0 的距离在法向量 (ω) 方向上的投影来表示, 即:



$$\text{距离} = \frac{|\omega^T(x - x_0)|}{\|\omega\|} = \frac{|\omega^T x + b|}{\|\omega\|}$$

6.2 试使用 LIBSVM, 在西瓜数据集 3.0 α 上分别用线性核和高斯核训练一个 SVM, 并比较其支持向量的差别.

表 4.5 西瓜数据集 3.0 α

编号	密度	含糖率	好瓜
1	0.697	0.460	是
2	0.774	0.376	是
3	0.634	0.264	是
4	0.608	0.318	是
5	0.556	0.215	是
6	0.403	0.237	是
7	0.481	0.149	是
8	0.437	0.211	是
9	0.666	0.091	否
10	0.243	0.267	否
11	0.245	0.057	否
12	0.343	0.099	否
13	0.639	0.161	否
14	0.657	0.198	否
15	0.360	0.370	否
16	0.593	0.042	否
17	0.719	0.103	否

在交叉验证评分上, 高斯核svm要高于线性核svm;

在模型复杂度上, 高斯核svm的支持向量的数目要多于线性核svm

高斯核svm更加复杂

6.4 试讨论线性判别分析与线性核支持向量机在何种条件下等价.

SVM 与 LDA 均可用于样本最优划分超平面的求解, 即法向向量, 线性核 SVM 的输入空间与特征空间相同, 二者取等条件为:

$$\mathbf{w} = \mathbf{S}_w^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i$$

6.5 试述高斯核 SVM 与 RBF 神经网络之间的联系.

SVM 的确与神经网络有密切联系: 若将隐层神经元数设置为训练样本数, 且每个训练样本对应一个神经元中心, 则以高斯径向基函数为激活函数的 RBF 网络恰与高斯核 SVM 的预测函数相同。

6.6 试析 SVM 对噪声敏感的原因.

SVM 的决策边界 (超平面) 是由支持向量所确定的, 即利用相对较少的数据特征来学得整个数据的特性。由于支持向量相对较少, 若噪声样本出现在其上, 容易对超平面的决策产生相对较大的影响, 所以 SVM 对噪声敏感。



THANK YOU
