

SVM视频录制

其他文档

SVM视频录制

1. 间隔和支持向量

- 1.1 什么是线性可分？
- 1.2 什么超平面？什么是最大间隔超平面？
- 1.3 什么是支持向量？
- 1.4 svm的最优化问题是什么？

2. 对偶问题

- 2.1 约束条件下的目标函数如何求解最优化问题？
- 2.2 怎么理解对偶问题
- 2.3 什么是对偶问题
- 2.4 KKT约束条件
- 2.4 求解svm的优化问题步骤

3. 软间隔

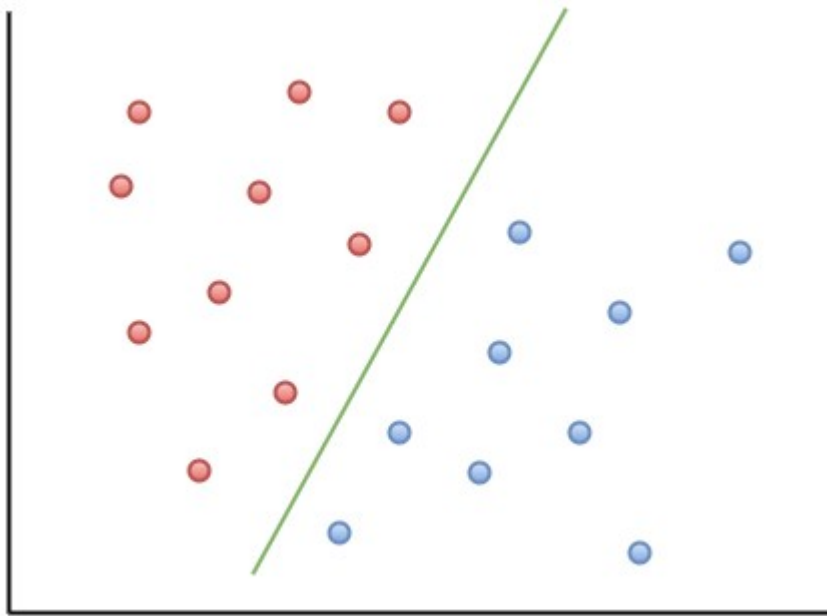
- 3.1 软间隔的提出是解决什么问题的？
- 3.2 软间隔后的线性svm的最优化问题是什么？

4. 核函数

- 4.1 线性不可分问题怎么解决？
- 4.2 什么是非线性SVM
- 4.3为什么要有核函数？
- 4.4 有了核函数，如何求解非线性SVM问题？
- 4.5 一些常用的核函数。

1. 间隔和支持向量

1.1 什么是线性可分？



上面的再二维界面上，两类点被一条直线完全分开，叫做线性可分。

严格的数学定义是：

D_0 和 D_1 是 n 维欧氏空间中的两个点集（点的集合）。如果存在 n 维向量 w 和实数 b ，使得所有属于 D_0 的点 x_i 都有 $w x_i + b > 0$ ，而对于所有属于 D_1 的点 x_j 则有 $w x_j + b < 0$ 。则我们称 D_0 和 D_1 线性可分。

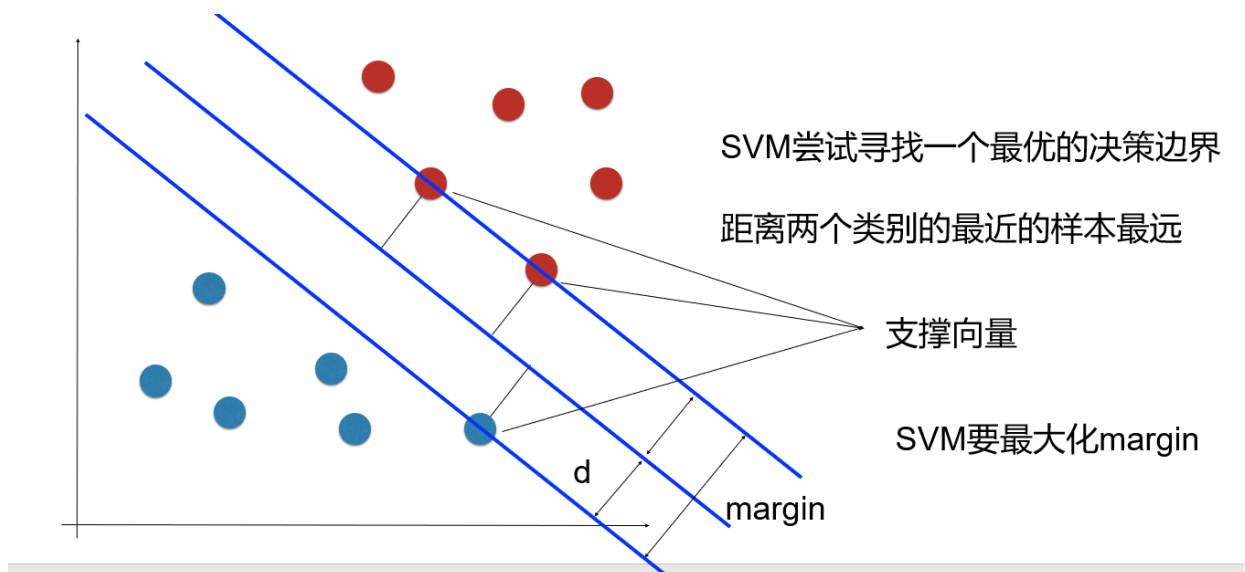
1.2 什么超平面？什么是最大间隔超平面？

上面提到的，将 D_0 和 D_1 完全正确地划分开的 $w x + b = 0$ ，就是一个超平面。

以最大间隔把两类样本分开的超平面，是最佳超平面！也称之为最大间隔超平面。

- 两类样本分别分隔在该超平面的两侧；
- 两侧距离超平面最近的样本点到超平面的距离被最大化了。

1.3 什么是支持向量？



在蓝色样本中存在一些距离我们的超平面最近的一些点，这些点叫做支持向量。

1.4 svm的最优化问题是什么？

首先我们想要最优化的是各类样本点到超平面的距离最远（其实也就是找到最大间隔超平面）。

然后任意一个超平面可以用下面这个线性方程来描述：

$$w^T x + b = 0$$

n维空间距离又是怎么算的呢？

我们看二维空间点 (x, y) 到直线的 $Ax + By + C = 0$ 的距离计算公式是：

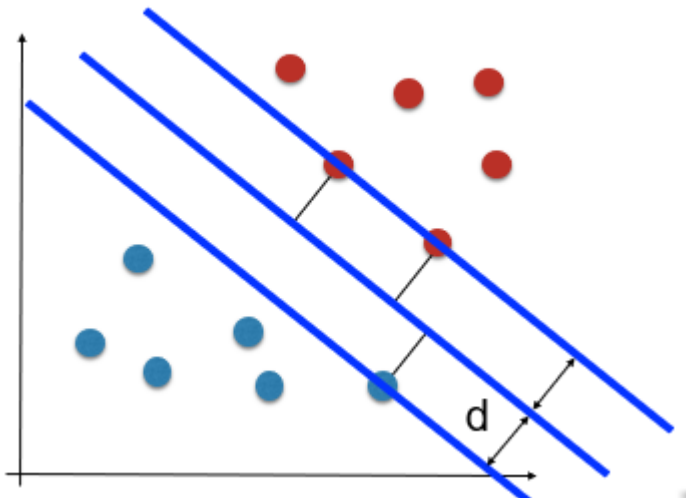
$$\frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$

那我们拓展到n维也是同样的，点 x 到直线 $w^T x + b = 0$ 的距离为：

$$\frac{|w^T x + b|}{\|w\|}$$

$$\text{其中}\|w\| = \sqrt{w_1^2 + \dots + w_d^2}$$

距离知道怎么计算了，再看这样一幅图：



根据支撑向量的定义，是样本中离超平面最近的点，所以所有的其他的红色点距离超平面的距离一定大于 d ，那么我们有这样一个公式：

$$\begin{cases} \frac{w^T x_i + b}{\|w\|} \geq d & y_i = 1 \\ \frac{w^T x_i + b}{\|w\|} \leq -d & y_i = -1 \end{cases}$$

稍作转化可以得到：

$$\begin{cases} \frac{w^T x_i + b}{\|w\|d} \geq 1 & y_i = 1 \\ \frac{w^T x_i + b}{\|w\|d} \leq -1 & y_i = -1 \end{cases}$$

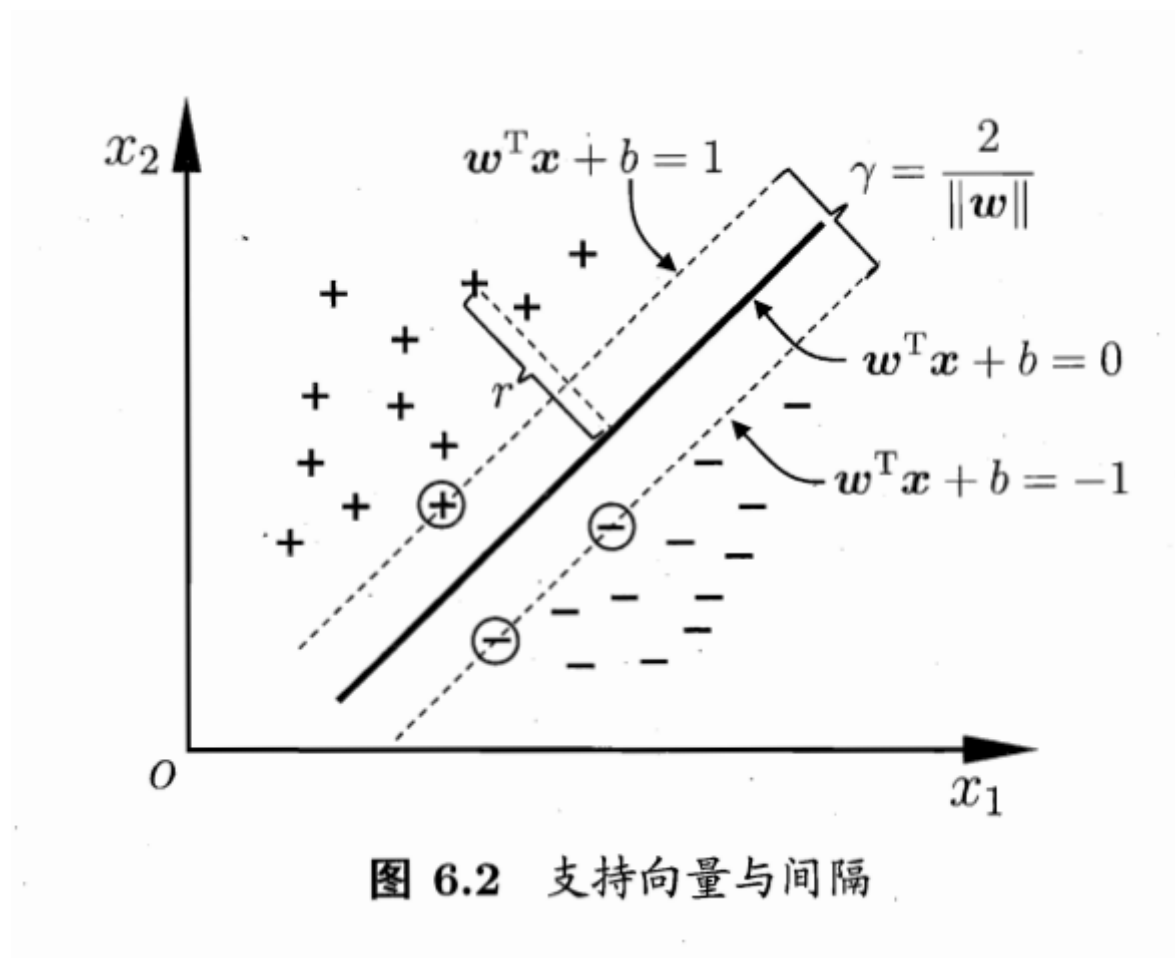
然后能够得到：

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

对于上面这个方程我们还能简写一下：

$$y_i(w^T x_i + b) \geq 1$$

所以我们可以得到上下两个超平面的方程为：



并且每个支撑向量到超平面的距离可以写为：

$$d = \frac{|w^T x_i + b|}{\|w\|}$$

我们要最大化这个距离也就是

$$\max \frac{|w^T x_i + b|}{\|w\|}$$

在样本点确定以后， $|w^T x_i + b|$ 是一个常数，所以这个式子就变成了：

$$\max \frac{1}{\|w\|}$$

也就是：

$$\min \|w\|$$

为了计算方便，我们取：

$$\min \frac{1}{2} \|w\|^2$$

所以得到最后的优化问题是：

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s. t.} \quad & y_i(w^T x_i + b) \geq 1 \end{aligned}$$

2. 对偶问题

2.1 约束条件下的目标函数如何求解最优化问题？

可以使用拉格朗日乘子法进行求解，将原本有约束的优化问题，转化为对拉格朗日函数的无约束优化问题了。

拉格朗日乘子法的定义（以三维空间为例）：

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y)$$

其中 λ 是拉格朗日乘子。

拉格朗日函数把原本的目标函数和其限制条件整合成了一个函数，这样子约束问题就不存在了，我们可以直接对该目标函数求解其极值。

2.2 怎么理解对偶问题

首先，原问题是：

$$\begin{aligned} \min_{w, b} \quad & \frac{\|w\|^2}{2} \\ \text{s. t.} \quad & g_i(w, b) = 1 - y_i(wx_i + b) \leq 0, \quad i = 1, 2, \dots, m \end{aligned}$$

拉格朗日乘子法转化后，变成了：

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \lambda_i [1 - y_i(wx_i + b)]$$

并且原问题变成了：

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

$$s. t. \quad \lambda_i \geq 0$$

我们首先直观的看一下，为什么转化后的下面的式子，可以替代上面的式子。

由于 $\lambda_i \geq 0$

当 $1 - y_i(wx_i + b) \geq 0$ 时， $\max_{\lambda} L(w, b, \lambda)$ 是无穷，无意义。

当 $1 - y_i(wx_i + b) \leq 0$ 时， $\max_{\lambda} L(w, b, \lambda)$ 是 $\frac{1}{2} \|w\|^2$

所以 $\min(\text{无穷大}, \frac{1}{2} \|w\|^2) = \frac{1}{2} \|w\|^2$

所以你看这个转化后的式子实际上和原来想表达的是的是一样的。

2.3 什么是对偶问题

我们简单点来讲，对偶问题实际上就是将

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

$$s. t. \quad \lambda_i \geq 0$$

变成了

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

$$s. t. \quad \lambda_i \geq 0$$

我们假设有一个函数 f 我们一定有：

$$\min \max f \geq \max \min f$$

也就是说，最大的里面挑出来的最小的，也要比最小的里面挑出来的最大的要大。

这个关系实际上就是弱对偶关系，那什么是强对偶关系呢？

$$\min \max f = \max \min f$$

我们上面实际上就是用了强对偶的关系。

为什么这里能够使用强对偶关系是不做要求的，但我们可以写出来，供大家扩展。

如果主问题是凸优化问题，也就是说当：

1. 拉格朗日函数中的 $f(x)$ 和 $g_j(x)$ 都是凸函数；
2. $h_i(x)$ 是仿射函数；
3. 主问题可行域中至少有一点使得不等式约束严格成立。即存在 x ，对所有 j ，均有 $g_j(x) < 0$ 。

1、2、3同时成立时，强对偶性成立。

2.4 KKT约束条件

首先我们看一下什么是kkt条件，以本节SVM为例：

1. $\frac{\partial L}{\partial w} = 0 \quad \frac{\partial L}{\partial b} = 0 \quad \frac{\partial L}{\partial \lambda} = 0$
2. $\lambda_i(1 - y_i(wx_i + b)) = 0$
3. $\lambda_i \geq 0$
4. $1 - y_i(wx_i + b) \leq 0$

什么时候KKT条件成立呢？

原问题具有强对偶性的**充要条件**是满足KKT条件

上面这个式子告诉了我们什么事？

直观的来讲就是，支撑向量满足 $1 - y_i(wx_i + b) = 0$ ，所以 $\lambda_i > 0$ 即可。

而其他向量， $1 - y_i(wx_i + b) < 0$ ， $\lambda_i = 0$

2.4 求解svm的优化问题步骤

我们已知svm优化的主问题是：

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

$$s.t. \quad g_i(w, b) = 1 - y_i(wx_i + b) \leq 0, \quad i = 1, 2, \dots, m$$

那么使用对偶算法求解线性可分的svm的步骤为：

步骤1：对主问题构造拉格朗日函数。

引入拉格朗日乘子 $\lambda_i \geq 0$ ，其中 $i = 1, 2, \dots, m$ ，得到拉格朗日函数：

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \lambda_i [1 - y_i(wx_i + b)]$$

步骤2：求拉格朗日函数对于 w ， b 的最小（还记得上面对偶问题的转换吗）。

我们再把上面对偶关系转化成了形式写一下：

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

$$s.t. \quad \lambda_i \geq 0$$

首先对 w 求偏导得到：

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^m \lambda_i x_i y_i = 0$$

然后再对 b 求偏导：

$$\frac{\partial L}{\partial b} = \sum_{i=1}^m \lambda_i y_i b = 0$$

最终得到结果：

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

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

我们将这个结果带回拉格朗日函数中可得：

$$\begin{aligned} L(w, b, \lambda) &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^m \lambda_i - \sum_{i=1}^m \lambda_i y_i \left(\left(\sum_{j=1}^m \lambda_j y_j x_j \right) \cdot x_i + b \right) \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^m \lambda_i - \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) - b \sum_{i=1}^m \lambda_i y_i \\ &= \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) \end{aligned}$$

实际上也就是：

$$\min_{w,b} L(w, b, \lambda) = \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j)$$

步骤3：求 $\min_{w,b} L(w, b, \lambda)$ 对 λ 的最大。

这还是上面所说的对偶问题，还记得什么是强对偶性吗。

所以现在我们求解变成了这样：

$$\max_{\lambda} \left[\sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) \right]$$

$$s. t. \sum_{i=1}^m \lambda_i y_i = 0$$

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

对于凸优化问题，我们习惯于求min而不是max，稍作转化问题变成了这样：

$$\min_{\lambda} \left[\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^m \lambda_i \right]$$

$$s. t. \sum_{i=1}^m \lambda_i y_i = 0$$

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

步骤4：由对偶问题求 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。

$$\text{设： } T(\lambda_1, \lambda_2, \dots, \lambda_m) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^m \lambda_i$$

问题就变成了求解 T 在约束条件下的最优解。

我们可以通过**SMO**算法求出这个最优解 λ^* 。

步骤5: 由 λ^* 求 w 。

由步骤1已知:

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

x_i 、 y_i 已知, λ_i^* 已由上一步求出, 将它们带入上式, 求 w 。

步骤6: 由 w 求 b 。

$\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$ 都已经求出来了。

因为 $\lambda_i (1 - y_i(wx_i + b)) = 0$; $i = 1, 2, \dots, m$ 是整体约束条件(KKT约束条件); 又因为对于所有支持向量(x_s, y_s), 都有 $1 - y_s(wx_s + b) = 0$, 因此, 所有大于0的 λ_k^* 所对应的(x_k, y_k)必然是支持向量。

这样我们就能确认样本点中哪些是支撑向量。

那么既然哪些 (x, y) 对是支持向量都已经清楚了, 理论上讲, 我们随便找一个支持向量(x_s, y_s), 把它和 w 带入: $y_s(wx_s + b) = 1$, 求出 b 即可。

$y_s(wx_s) + y_s b = 1$, 两边乘以 y_s 。

$y_s^2(wx_s) + y_s^2 b = y_s$, 因为 $y_s^2 = 1$, 所以: $b = y_s - wx_s$ 。

为了更加鲁棒, 我们可以求所有支持向量的均值:

$$b = \frac{1}{|S|} \sum_{s \in S} (y_s - wx_s)。$$

步骤7: 求最终结果。

现在 w 和 b 都求出来了

我们就能构造出最大分割超平面: $wx + b = 0$ 。

构造分类决策函数: $f(x) = \text{sign}(wx + b)$ 。

其中, $\text{sign}(\cdot)$ 全称为 *SignumFunction*。其定义为:

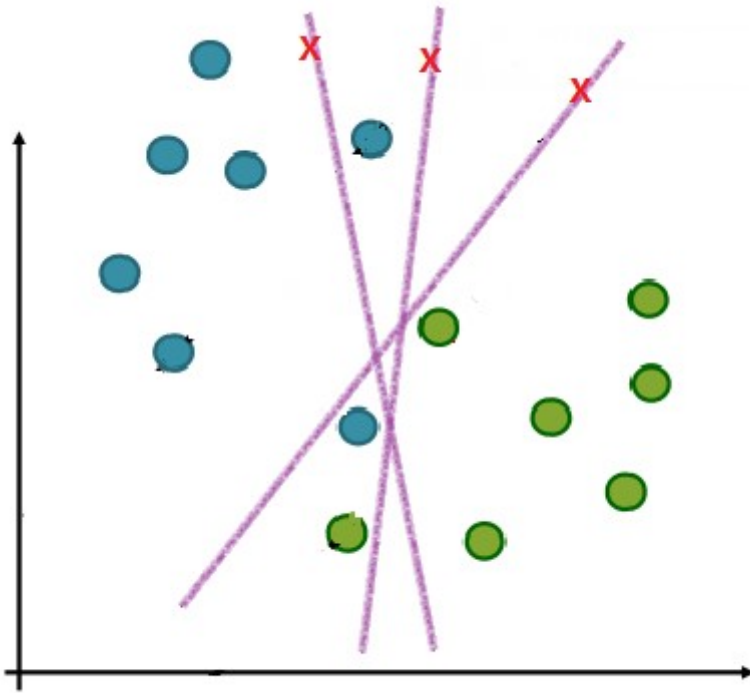
$$\text{sign}(x) = \begin{cases} -1 & : x < 0 \\ 0 & : x = 0 \\ 1 & : x > 0 \end{cases}$$

将新的样本点带入到决策函数中, 可以得到样本的分类。

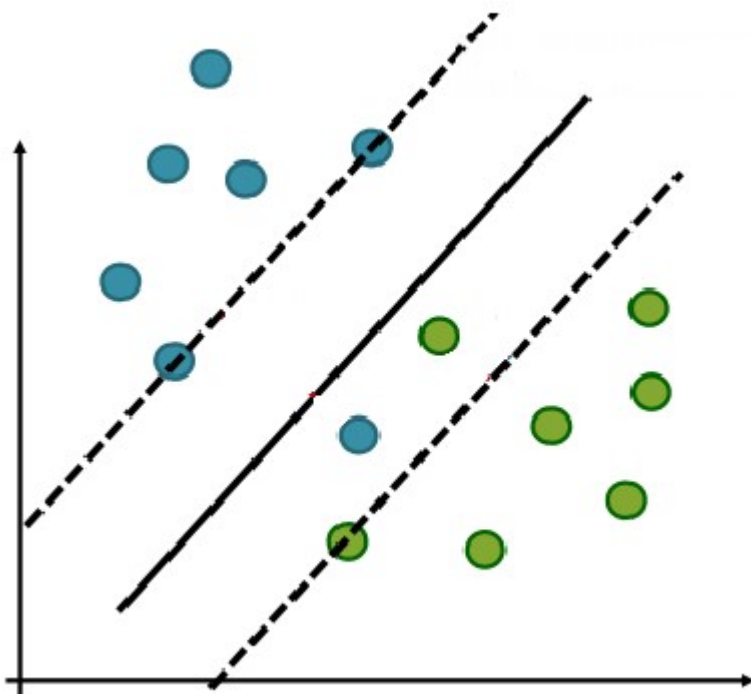
3. 软间隔

3.1 软间隔的提出是解决什么问题的?

在实际的应用中，完全线性可分的样本是很少的，如果遇到了不能够完全线性可分的样本，比如下面这个：



我们应该怎么办？所以提出了软间隔，允许个别样本点出现在间隔带里面，比如：

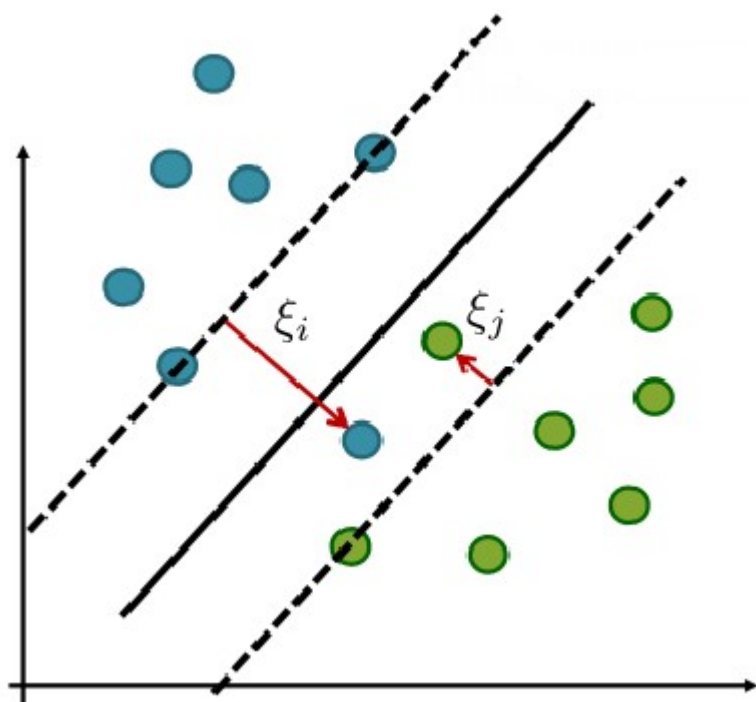


这样就允许了部分样本点不满足约束条件：

$$y_i(wx_i + b) \geq 1$$

为了度量这个间隔“软”到何种程度，我们针对每一个样本 (x_i, y_i) ，引入一个松弛变量 ξ_i ，令 $\xi_i \geq 0$ ，且 $y_i(wx_i + b) \geq 1 - \xi_i$ 。

对应到图上就是：



3.2 软间隔后的线性svm的最优化问题是什么？

优化目标

从原来的线性可分的优化目标

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

$$s.t. \quad 1 - y_i(wx_i + b) \leq 0, \quad i = 1, 2, \dots, m$$

变成了：

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

$$s.t. \quad y_i(wx_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m; \quad \xi_i \geq 0, \quad i = 1, 2, \dots, m$$

其中 C 是一个大于0的常数，若 C 为无穷大，则 ξ_i 必然为无穷小，否则将无法最小化主问题。如此一来，线性 SVM 就又变成了线性可分 SVM。

当 C 为有限值的时候，才能允许**部分样本**不遵守约束条件 $1 - y_i(wx_i + b) \leq 0$ 。

我们将优化目标整理成和线性可分的形式一样即为：

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

$$s.t. \quad 1 - \xi_i - y_i(wx_i + b) \leq 0, \quad i = 1, 2, \dots, m; \quad -\xi_i \leq 0, \quad i = 1, 2, \dots, m$$

我们开始和上面一样使用对偶算法求解上面的最优化问题：

步骤1：构造拉格朗日函数

$$L(w, b, \xi, \lambda, \mu) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i + \sum_{i=1}^m \lambda_i [1 - \xi_i - y_i(wx_i + b)] + \sum_{i=1}^m (-\mu_i \xi_i)$$

$$\lambda_i \geq 0, \mu_i \geq 0$$

其中 λ_i 和 μ_i 是拉格朗日乘子，而 w 、 b 和 ξ_i 是主问题参数。

根据主问题的对偶性，主问题的对偶问题是：

$$\max_{\lambda, \mu} \min_{w, b, \xi} L(w, b, \xi, \lambda, \mu)$$

步骤2：最小化拉格朗日函数

首先对 w 、 b 和 ξ 最小化 $L(w, b, \xi, \lambda, \mu)$ ——分别对 w 、 b 和 ξ_i 求偏导，然后令导数为0，得出如下关系：

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

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

$$C = \lambda_i + \mu_i$$

将这些关系带入线性 SVM 主问题的拉格朗日函数，得到：

$$\min_{w, b, \xi} L(w, b, \xi, \lambda, \mu) = \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j)$$

步骤3：求 $\min_{w, b} L(w, b, \xi, \lambda, \mu)$ 对 λ 的最大。

因为上面最小化的结果中只有 λ 而没有 μ ，所以现在只需要最大化 λ 就好：

$$\max_{\lambda, \mu} \min_{w, b, \xi} L(w, b, \xi, \lambda, \mu) = \max_{\lambda} \left(\sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) \right)$$

$$s. t. \quad \sum_{i=1}^m \lambda_i y_i = 0; \quad C - \lambda_i - \mu_i = 0; \quad \lambda_i \geq 0; \quad \mu_i \geq 0; \quad i = 1, 2, \dots, m$$

步骤4：使用SMO算法求解

我们同样可以和之前一样，将最大化问题转化为最小化问题：

$$\max_{\lambda} \left(\sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) \right) = \min \left(\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^m \lambda_i \right)$$

$$s. t. \quad \sum_{i=1}^m \lambda_i y_i = 0; \quad 0 \leq \lambda_i \leq C; \quad i = 1, 2, \dots, m$$

然后我们可以看到，实际上差别只在约束条件上，所以我们同样可以使用SMO算法进行求解出拉个朗日乘子 λ^* 。

步骤5：求解 w 和 b

由 $w = \sum_{i=1}^m \lambda_i y_i x_i$ 求出 w 。

因为最终要求得的超平面满足 $w \cdot x + b = 0$ ，这一点是和线性可分 SVM 的超平面一样的，因此求解 b 的过程也可以照搬：

$$b = \frac{1}{|S|} \sum_{s \in S} (y_s - w x_s)$$

其中 S 是支持向量的集合。

在这里面会存在一个问题，就是到底那部分在间隔中的样本点，是不是支撑向量？

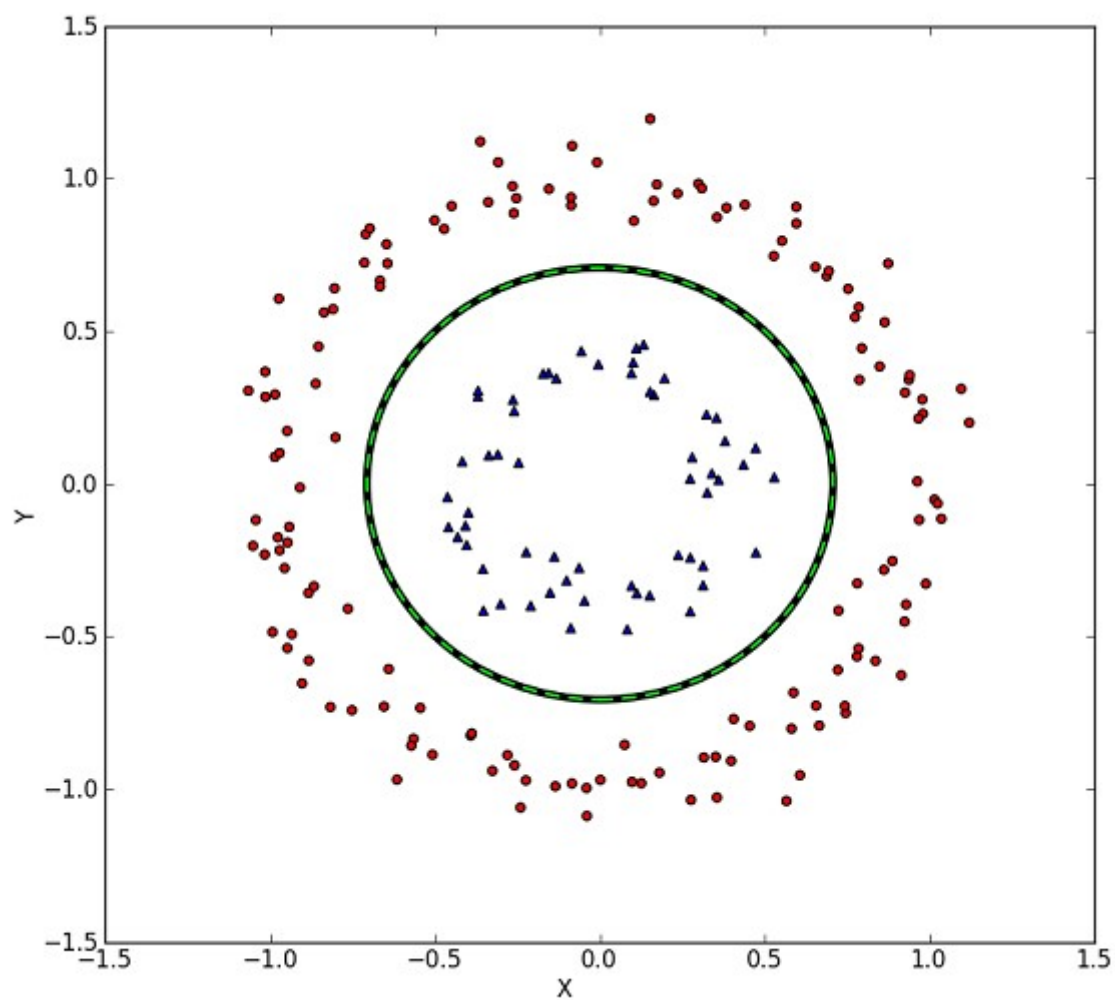
我们可以看到 $w = \sum_{i=1}^m \lambda_i y_i x_i$ ，因此对于所有的 $\lambda_i > 0$ 的点都能够影响到我们的超平面，因此都是支撑向量。

4. 核函数

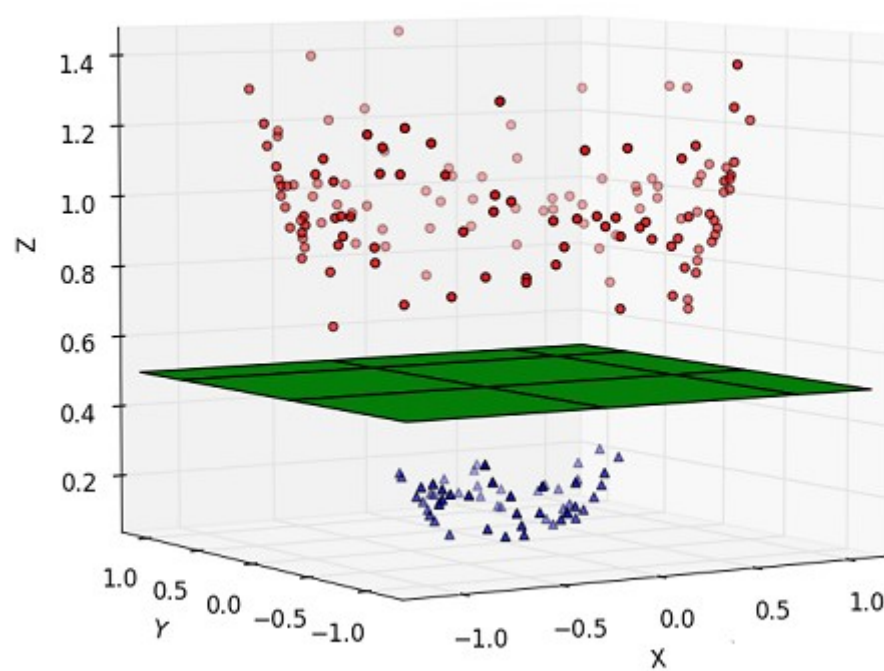
4.1 线性不可分问题怎么解决？

对svm来讲，最好的当然是样本完全线性可分，就算差一点不是完全的我们也希望绝大部分样本点线性可分。

但是我们可能碰到一种情况，样本点不是线性可分的，比如：



这种情况的一种解决办法就是，将二维线性不可分样本，映射到高维空间去，让样本点在高维空间线性可分，比如变成这样：



4.2 什么是非线性SVM

对于在有限维度向量空间中线性不可分的样本，我们将其映射到更高维度的向量空间里，再通过间隔最大化的方式，学习得到支持向量机，就是**非线性 SVM**。

我们将样本映射到的这个更高维度的新空间叫做**特征空间**。

我们现在用 x 来表示原来的样本点，用 $\phi(x)$ 表示 x 映射到特征空间之后的新向量。

那么分隔超平面可以表示为： $f(x) = w\phi(x) + b$ 。

由上面的线性 SVM 的对偶问题，此处非线性 SVM 的对偶问题就变成了：

$$\min(\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j \phi(x_i) \cdot \phi(x_j) - \sum_{i=1}^m \lambda_i)$$

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

$$0 \leq \lambda_i \leq C, \quad i = 1, 2, \dots, m$$

可以看到，和线性 SVM 唯一的不同就是：之前的 x_i 与 x_j 的内积（点乘）变成了 $\phi(x_i)$ 与 $\phi(x_j)$ 的内积。

4.3 为什么要有核函数？

我们可以看到，对于非线性函数来说，和之前不同的主要是 $\phi(x_i)$ 与 $\phi(x_j)$ 的内积。而由于是低维空间映射到高维空间，维度可能会很大，所以如果将全部样本的点乘全部计算好，这样的计算量太大了。

针对这个问题，我们的解决方式是**先不算**，用一个函数先代替，我们用

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

这样子函数 $k(x_i, x_j)$ 称之为核函数

这样子大家可能还体会不到核函数的好处，那我们继续看下面的一个例子：假设我们现在有一个多项式核函数：

$$k(x, y) = (x \cdot y + 1)^2$$

对于这个核函数，如果带进样本的话，是这样的：

$$k(x, y) = (\sum_{i=1}^n (x_i \cdot y_i) + 1)^2$$

这个式子展开式是什么样子的？

$$\sum_{i=1}^n (x_i^2) (y_i^2) + \sum_{i=2}^n \sum_{j=1}^{i-1} (\sqrt{2}x_i x_j) (\sqrt{2}y_i y_j) + \sum_{i=1}^n (\sqrt{2}x_i) (\sqrt{2}y_i) + 1$$

我们需要先把我们的向量映射成：

$$x' = (x_n^2, \dots, x_1^2, \sqrt{2}x_n x_{n-1}, \dots, \sqrt{2}x_n, \dots, \sqrt{2}x_1, 1)$$

然后再进行计算，可见这个映射不管是计算量还是存储量都是非常巨大的。但是有了核函数，就不需要作这样的映射，直接使用原样本维度的点进行计算即可。

4.4 有了核函数，如何求解非线性SVM问题？

在有了核函数之后，非线性问题重新转变成了线性问题，和之前求解过程一样，先根据对偶函数求解 λ ，然后根据 λ 求解 w ，再根据支撑向量求解 b 即可。

4.5 一些常用的核函数。

线性核函数

$$k(x_i, x_j) = x_i^T x_j$$

使用时无须指定参数，它直接计算两个输入向量的内积。经过线性核函数转换的样本，特征空间与输入空间重合，相当于并没有将样本映射到更高维度的空间里去。

很显然这是最简单的核函数，实际训练、使用 SVM 的时候，在不知道用什么核的情况下，可以先试试线性核的效果。

多项式核 (Polynomial Kernel)

$$k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad \gamma > 0, d \geq 1$$

需要指定三个参数： γ 、 r 和 d 。

这是一个不平稳的核，适用于数据做了归一化的情况。

RBF 核 (Radial Basis Function Kernel)

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0$$

RBF 核又名高斯核 (Gaussian Kernel)，是一个核函数家族。它会将输入空间的样本以非线性的方式映射到更高维度的空间（特征空间）里去，因此它可以处理类标签和样本属性之间是非线性关系的状况。

它有一个参数： γ ，这个参数的设置非常关键！

如果设置过大，则整个 RBF 核会向线性核方向退化，向更高维度非线性投影的能力就会减弱；但如果设置过小，则会使得样本中噪音的影响加大，从而干扰最终 SVM 的有效性。

不过相对于多项式核的3个参数，RBF 核只有一个参数需要调，还是相对简单的。

当线性核效果不是很好时，可以用 RBF 试试。或者，很多情况下可以直接使用 RBF。

在具体应用核函数时，最好针对具体问题参照前人的经验。