支持向量机（SVM）

标签（空格分隔）： 监督学习

@author : duanxxnj@163.com
@time : 2016-07-31

二、软间隔最大化分类器

在这篇文章之前所有的讨论中，假设，所面对的数据在特征空间$\phi(x)$ ϕ(x)中是线性可分的，那么，最终得到的SVM可以在输入空间X X中得到一个准确的判别面，尽管其对应的判别面可能是一个非线性判别面。

然而，在现实的数据中，数据是线性不可分的（或者近似线性可分）。换一种说法就是：类条件分布存在重叠。如果这样，上面讲的算法最终只会有两种结果：1、决策面会出现严重的震荡，而无法收敛；2、即便收敛到了一个非线性的决策面，其决策面也会十分的复杂，泛化能力低。

因此，我们就需要对原始的SVM算法做一定的改动，让它对数据的要求，从线性可分扩展到线性不可分。即：允许部分训练数据存在分类错误的情况，以得到一个相对比较好的决策面。 这个改动后的SVM一般被称为 **软间隔最大化分类器** ， 而与之相对应的，前面论述的只可以解决线性可分数据的SVM，称为 **硬间隔最大化分类器** 。

2.1 软间隔最大化分类器的两个问题

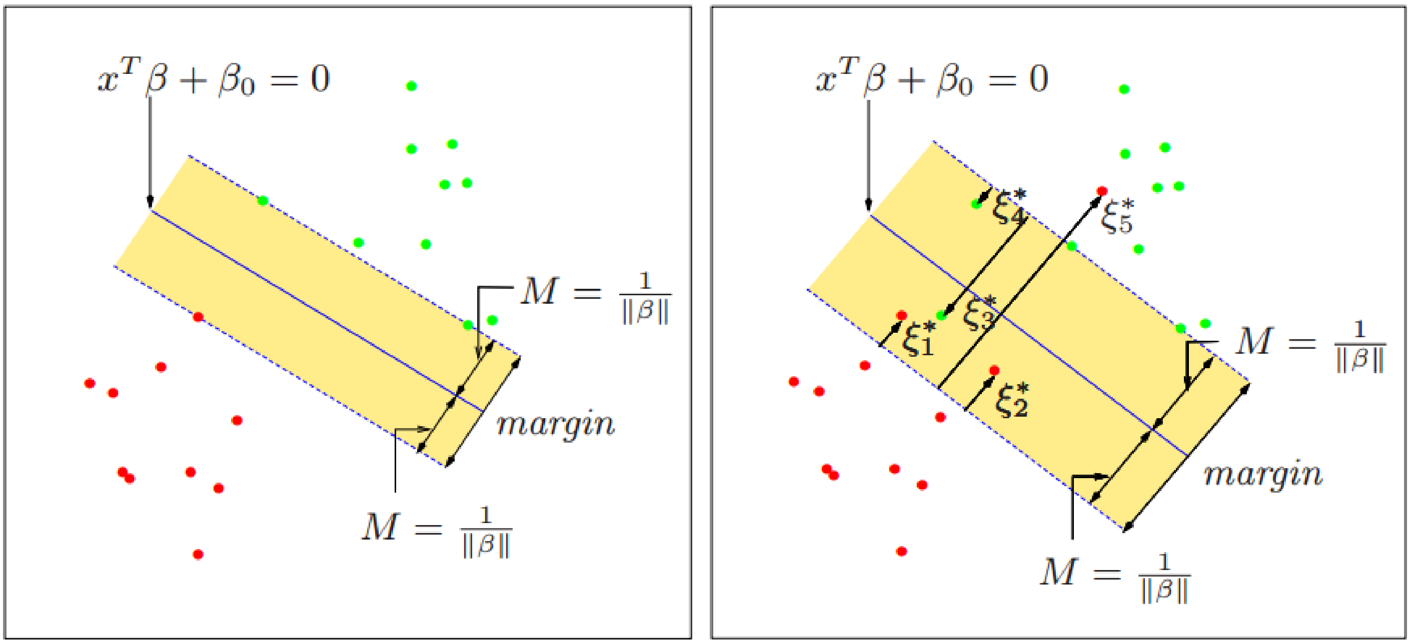
那么，现在就有两个问题摆在我们面前：

1.在SVM中，对于非线性可分的数据而言，哪些样本点是存在分类错误的样本点？
2.对于分类错误的点，**软间隔最大化分类器** 如何处理以得到相对比较好的决策面？

两个问题，就是软间隔最大化分类器的核心问题，回答了这两个问题，掌握了软间隔最大化分类器。

哪些是分类错误的样本点

首先必须要知道，面对一个非线性可分的数据集，哪些样本点是分类错误的样本点？可以先看下面这幅图：



图中，左边那幅图对应的是数据线性可分时，SVM的状态。此时有决策面：$x^T\beta + \beta_0 = 0$，也就是之前讨论的 $x^Tw + b = 0$，只是写法不同而已。根据前面“最小间隔最大化”小章节的讨论可知：最小间隔为$M = 1/||\beta||$，也就是之前的表达式：$r = 1/||w||$。这种状态，是所有数据点都分类正确，且最小间隔最大分类的状态。

图中，右边那幅图对应的就是数据线性不可分时，SVM的状态。此时决策面依然是：$x^T\beta + \beta_0 = 0$，最小间隔仍然为$M = 1/||\beta||$。不过，右图中$\xi_i^*; i = 1, 2...5$ 就是非线性可分数据在SVM中分类错误的点。

那么总结一下，在非线性可分的情况下，样本点的状态一共有下面五种：
1.非支持向量，即是在 $margin$ 两侧，被正确分类的样本点。这些样本点对于SVM的决策面没有贡献。
2.在$margin$边界上面的点，这个是支持向量，且被正确分类。
3.在$margin$内部，被正确分类的点，比如$\xi_1^*, \xi_2^*, \xi_4^*$，这些也是支持向量。
4.在决策面上的点，这些点可以拒绝判断，也可以直接认为是分类正确的点。
5.真正的被分类错误的点：ξ_3^*, ξ_5^*，这些点直接走到了决策面的另一边去了。

很容易发现，相对于数据线性可分的情况而言，非线性可分数据的状态多了三种情况，也就是上面五种情况的后三种，线性可分的数据只存在前面两种情况。

而后面三种情况的样本点的间隔，都比最小间隔 r 也就是图中的 M 要小一些，所以，后面这三种情况，就是SVM认为的错误的样本点（即便情况3、情况4样本被分类正确了），**软间隔最大化分类器** 就是要针对后面这三种情况来做文章。

分类错误的样本点的处理:软间隔变量

前面提到了，后面三种情况的样本点的间隔，都比最小间隔 r 也就是图中的 M 要小一些。这里首先要考虑的问题就是，如何用公式来表达后面这三种情况？这里就引入了一个最小间隔 r 的比例因子：软间隔变量 ξ。ξ是最小间隔 r 的比例系数，或者叫做倍数。下面先分别讨论如何将软间隔变量 ξ 应用到上面的后三种情况中：

- 对于前面第三种情况：在$margin$内部，被正确分类的点，比如$\xi_1^*, \xi_2^*, \xi_4^*$。这些点的间隔为：

$$\xi_i^* = \xi_i * r; \quad 0 < \xi_i < 1$$

- 对于前面第四种情况：在决策面上的点，这些点的间隔为：

$$\xi_i^* = \xi_i * r; \quad \xi_i = 1$$

ξi∗=ξi∗r; ξi=1

- 对于前面第五种情况：真正的被分类错误的点：ξ_3^*, ξ_5^* ξ3∗,ξ5∗，这些直接走到了决策面的另一边去了的点，这些点的间隔为：

$$\xi_i^* = \xi_i * r; \quad \xi_i > 1$$

ξi∗=ξi∗r; ξi>1

为了将前面的第一种情况：非支持向量，即是在 $margin$margin 两侧，被正确分类的样本点。和第二种情况：在 $margin$margin边界上面的点，这个是支持向量，且被正确分类。也统一的用软间隔变量 ξξ来表示，这里将SVM的优化公式重写为下面这个形式：

$$\max_{w,b} r$$
$$\text{st：} \frac{t_i\{w^T \phi(x_i) + b\}}{||w||} \ge r(1-\xi_i); \quad i = 1,2,\ldots N$$
$$\xi_i \ge 0; i = 1,2,\ldots N$$

maxw,brst：ti{wTϕ(xi)+b}||w||≥r(1−ξi); i=1,2,...Nξi≥0;i=1,2,...N

而前面第一种情况和第二种情况所对应的 $\xi_i = 0$ξi=0。

和"最小间隔最大化"章节中讨论的一样，ww 和 bb 的 **比值** 不变的情况下，ww是可以任意取值的。这样的话，为了便于计算，我们就取：

$$||w|| = \frac{1}{r}$$

||w||=1r

那么，上面的式子又可以改写为：

$$\max_{w,b} \frac{1}{||w||}$$
$$\text{st：} t_i\{w^T \phi(x_i) + b\} \ge 1 - \xi_i; i = 1,2,\ldots N$$
$$\xi_i \ge 0; i = 1,2,\ldots N$$

maxw,b1||w||st：ti{wTϕ(xi)+b}≥1−ξi;i=1,2,...Nξi≥0;i=1,2,...N

可以看到，上面的式子是最大化 $1/||w||$1/||w||，很容易知道，其等价于 最小化 $||w||^2$||w||2，那么最小间隔最大化，最终就可以变为下面这个式子：

$$\min_{w,b} ||w||^2$$
$$\text{st：} t_i\{w^T \phi(x_i) + b\} \ge 1 - \xi_i; i = 1,2,\ldots N$$
$$\xi_i \ge 0; i = 1,2,\ldots N$$

minw,b||w||2st：ti{wTϕ(xi)+b}≥1−ξi;i=1,2,...Nξi≥0;i=1,2,...N

这里，最小间隔为：

$$r = \frac{1}{||w||}$$

r=1||w||

2.2 软间隔最大化分类器 的优化公式

上面提到的第五种情况，真正的被分类错误的点：ξ_3^*, ξ_5^*，这些直接走到了决策面的另一边去了的点，所对应的 $\xi_i > 1$。如果说我们知道 $\sum_{i=1}^{N} \xi_i \le K$，那么，就可以说，这个 **软间隔最大化分类器** 的训练错误样本数，最多为 K 个。让训练误差尽量的小也是我们的一个优化目标，这样，**软间隔最大化分类器** 就成了一个多目标的优化问题：

$$\begin{gathered}\min_{w,b} \frac{1}{2}||w||^2 \\ \min_{\xi} \sum_{i=1}^{N} \xi_i \\ \text{st：} t_i\{w^T \phi(x_i) + b\} \ge 1 - \xi_i ; i = 1, 2, \ldots N \\ \xi_i \ge 0; i = 1, 2, \ldots N\end{gathered}$$

minw,b12||w||2minξ∑i=1Nξist：ti{wTϕ(xi)+b}≥1−ξi;i=1,2,...Nξi≥0;i=1,2,...N

由于上面两个优化目标的限制条件是一样的，可以将两个优化目标合成一个优化目标：

$$\begin{gathered}\min_{w,b,\xi} \{\frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i\} \\ \text{st：} t_i\{w^T \phi(x_i) + b\} \ge 1 - \xi_i ; i = 1, 2, \ldots N \\ \xi_i \ge 0; i = 1, 2, \ldots N\end{gathered}$$

minw,b,ξ{12||w||2+C∑i=1Nξi}st：ti{wTϕ(xi)+b}≥1−ξi;i=1,2,...Nξi≥0;i=1,2,...N

这里的参数 C 是损失参数（cost parameter），用于权衡 $||w||^2$ 和 $\sum_{i=1}^{N} \xi_i$。一般我们将 $||w||^2$ 视为模型的复杂度，而将 $\sum_{i=1}^{N} \xi_i$ 视为模型的错误率，损失参数 C 就是在模型复杂度和模型的错误率之间找一个平衡，这个值一般作为模型的一个输入变量，需要人工设置：

- C 越小，模型越偏向于最小化 $||w||^2$，而忽略 $\sum_{i=1}^{N} \xi_i$，这样模型非常简单，但模型的错误率相对较高。
- C 越大，模型越偏向于最小化 $\sum_{i=1}^{N} \xi_i$，而忽略 $||w||^2$，这样模型的错误率相对较低，但模型相对较为复杂。

> 需要强调：ξ_i 是对于SVM而言，分类错误的样本点 x_i 的分类错误间隔相对于最小间隔 r 的比例。我们不可能让错误的比例无限的增长，那样分类错误就太大了。所以，我们需要对错误间隔进行限制，而限制的手段，就是为总的错误间隔比例加一上限：$\sum \xi_i \le constant$。这是**软间隔最大化分类器** 优化公式的另一种解释。

2.3 软间隔最大化分类器 的对偶问题

这里的步骤和 1.5 拉格朗日对偶性 几乎雷同，首先将 **软间隔最大化分类器** 的优化公式转换为拉个朗日函数的形式，其原始的优化公式为：

$$\begin{gathered}\min_{w,b,\xi} \{\frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i\} \\ \text{st：} t_i\{w^T \phi(x_i) + b\} \ge 1 - \xi_i ; i = 1, 2, \ldots N \\ \xi_i \ge 0; i = 1, 2, \ldots N\end{gathered}$$

minw,b,ξ{12||w||2+C∑i=1Nξi}st：ti{wTϕ(xi)+b}≥1−ξi;i=1,2,...Nξi≥0;i=1,2,...N

拉个朗日函数为：

$$L(w, b, \xi, \alpha, \mu) = \underbrace{\frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i}_{\text{优化目标}} \underbrace{- \sum_{i=1}^{N} \alpha_i\{t_i\{w^T \phi(x_i) + b\} - 1 + \xi_i\} - \sum_{i=1}^{N} \mu_i \xi_n}_{\text{不等式约束}}$$

L(w,b,ξ,α,μ)=12||w||2+C∑i=1Nξi⏟优化目标−∑i=1Nαi{ti{wTϕ(xi)+b}−1+ξi}−∑i=1Nμiξn⏟不等式约束

其中，α_i, μ_i 为拉格朗日乘子。将 $L(w,b,\xi,\alpha,\mu)$ 分别对 w,b,ξ 求偏导:

$$\begin{aligned} \nabla_w L(w,b,\xi,\alpha,\mu) &= w - \sum_{i=1}^{N} \alpha_i t_i \phi(x_i) = 0 \\ \nabla_b L(w,b,\xi,\alpha,\mu) &= \sum_{i=1}^{N} \alpha_i t_i = 0 \\ \nabla_{\xi_i} L(w,b,\xi,\alpha,\mu) &= C - \alpha_i - \beta_i = 0; i = 1,2,..N \end{aligned}$$

∇wL(w,b,ξ,α,μ)=w−∑i=1Nαitiϕ(xi)=0∇bL(w,b,ξ,α,μ)=∑i=1Nαiti=0∇ξiL(w,b,ξ,α,μ)=C−αi−βi=0;i=1,2,..N

就可以得到:

$$\begin{aligned} w &= \sum_{i=1}^{N} \alpha_i t_i \phi(x_i) \\ \sum_{i=1}^{N} \alpha_i t_i &= 0 \\ C - \alpha_i - \beta_i &= 0; i = 1,2..N \end{aligned}$$

w=∑i=1Nαitiϕ(xi)∑i=1Nαiti=0C−αi−βi=0;i=1,2..N

将上面三个式子带入到 $L(w,b,\xi,\alpha,\mu)$ 中，和之前 1.6 最小间隔最大化求解 中的推导一样，最终可以得到:

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\alpha_i\alpha_j t_i t_j < \phi(x_i), \phi(x_j) >\} + \sum_{i=i}^{N} \alpha_i$$

minw,b,ξL(w,b,ξ,α,μ)=−12∑i=1N∑j=1N{αiαjtitj<ϕ(xi),ϕ(xj)>}+∑i=iNαi

这里的 $< \phi(x_i), \phi(x_j) >$ 是 $\phi(x_i)$ 和 $\phi(x_j)$ 的内积。

再对 $\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求极大化，即可得到 **软间隔最大化分类器** 的对偶问题:

$$\max_{\alpha}\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\alpha_i\alpha_j t_i t_j < \phi(x_i), \phi(x_j) >\} + \sum_{i=i}^{N} \alpha_i\}$$

maxα{−12∑i=1N∑j=1N{αiαjtitj<ϕ(xi),ϕ(xj)>}+∑i=iNαi}

$$\begin{aligned} s.t. \quad \sum_{i=1}^{N} \alpha_i t_i &= 0 \\ C - \alpha_i - \mu_i &= 0 \\ \alpha_i &\geq 0 \\ \mu_i &\geq 0; i = 1,2,...N \end{aligned}$$

s.t. ∑i=1Nαiti=0C−αi−μi=0αi≥0μi≥0;i=1,2,...N

这个就是原问题的对偶问题，当然了，可以将这个对偶问题的目标函数的符号换一下,,让它成为一个最小化的问题，并将后面三个式子做一个合并:

$$\min_{\alpha}\{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\alpha_i\alpha_j t_i t_j < \phi(x_i), \phi(x_j) >\} - \sum_{i=i}^{N} \alpha_i\}$$

minα{12∑i=1N∑j=1N{αiαjtitj<ϕ(xi),ϕ(xj)>}−∑i=iNαi}

$$\begin{aligned} s.t. \quad \sum_{i=1}^{N} \alpha_i t_i &= 0 \\ 0 \leq \alpha_i &\leq C; i = 1,2,...N \end{aligned}$$

s.t. ∑i=1Nαiti=00≤αi≤C;i=1,2,...N

这就是 **软间隔最大化分类器** 的对偶问题，和前面相比，这里的对偶问题只是限制条件 $0 \le \alpha_i \le C$ 和前面不同，其他的是一模一样的。

2.4 软间隔最大化分类器 求解

很显然，得到了**软间隔最大化分类器** 的对偶形式之后，其优化求解出来的就是最优解 α^*。这里再次说明：假设 α^* 我们已经通过某种算法得到了，至于这个算法是什么后面的章节会详细的说明，这里不同担心。

仿照 1.5 拉格朗日对偶性 中的公式推导， 假设 $x^*, \alpha^*, \xi^*, \alpha^*, \mu^*$ 为最优解，可以很容易推出 **软间隔最大化分类器** 拉格朗日优化的KKT条件为：

$$
\begin{aligned}
\nabla_w L(x^*,\alpha^*,\xi^*,\alpha^*,\mu^*) &= w^* - \sum_{i=1}^{N} \alpha_i^* t_i \phi(x_i) = 0 \\
\nabla_b L(x^*,\alpha^*,\xi^*,\alpha^*,\mu^*) &= \sum_{i=1}^{N} \alpha_i^* t_i = 0 \\
\nabla_\xi L(x^*,\alpha^*,\xi^*,\alpha^*,\mu^*) &= C - \alpha_i - \mu_i = 0 \\
\alpha_i^*(t_i\{w^{*T}\phi(x_i) + b^*\} - 1 + \xi_i^*) &= 0; i = 1,2,\dots N \\
t_i\{w^{*T}\phi(x_i) + b^*\} - 1 + \xi_i^* &\ge 0; i = 1,2,\dots N \\
\alpha_i^* &\ge 0; i = 1,2,\dots N \\
\mu_i^* \xi_i^* &= 0; i = 1,2,\dots N \\
\xi_i^* &\ge 0; i = 1,2,\dots N \\
\mu_i^* &\ge 0; i = 1,2,\dots N
\end{aligned}
$$

∇wL(x∗,α∗,ξ∗,α∗,μ∗)=w∗−∑i=1Nαi∗tiϕ(xi)=0∇bL(x∗,α∗,ξ∗,α∗,μ∗)=∑i=1Nαi∗ti=0∇ξL(x∗,α∗,ξ∗,α∗,μ∗)=C−αi−μi=0αi∗(ti{w∗Tϕ(xi)+b∗}−1+ξi∗)=0;i=1,2,...Nti{w∗Tϕ(xi)+b∗}−1+ξi∗≥0;i=1,2,...Nαi∗≥0;i=1,2,...Nμi∗ξi∗=0;i=1,2,...Nξi∗≥0;i=1,2,...Nμi∗≥0;i=1,2,...N

在假设知道 α^* 的情况下，由 KKT 条件的第一个公式：$\nabla_w L(x^*,\alpha^*,\xi^*,\alpha^*,\mu^*) = w^* - \sum_{i=1}^{N} \alpha_i^* t_i \phi(x_i) = 0$ 可以知道，w的最优解为：

$$w^* = \sum_{i=1}^{N} \alpha_i^* t_i \phi(x_i)$$

w∗=∑i=1Nαi∗tiϕ(xi)

再找到一个 $0 < \alpha_j < C$，其所对应的样本点满足：$t_i\{w^{*T}\phi(x_i) + b^*\} - 1 = 0$，那么，就可以和前面的推导一样，得到：

$$b^* = \frac{1}{t_i} - w^{*T}\phi(x_i)$$

b∗=1ti−w∗Tϕ(xi)

同时，需要注意：$t_i^2 = 1$，并带入 w^{*T}，就可以将上面这个式子重写为：

$$b^* = t_i - \sum_{j=1}^{N} \alpha_j^* t_j < \phi(x_j), \phi(x_i) >$$

b∗=ti−∑j=1Nαj∗tj<ϕ(xj),ϕ(xi)>

当然，为了稳妥起见，很多时候，我们会将所有的支持向量 $x_i \in S$ 对应的 b_i^* 都求出来，然后用其均值，作为最终的 b^*。这里 S 是支持向量的集合，也就是 $\alpha_i^* > 0$ 所对应的点集：

$$b^* = \frac{1}{N_S}\sum_{i=1}^{N_S} b_i^*$$

b∗=1Ns∑i=1Nsbi∗

这样，就可以求得最终的决策超平面为：

$$\sum_{i=1}^{N}\{\alpha_i^* t_i < \phi(x_i), \phi(x) >\} + b^* = 0$$

∑i=1N{αi∗ti<ϕ(xi),ϕ(x)>}+b∗=0

分类决策函数可以写为：

$$f(x) = \mathrm{sign}\{\sum_{i=1}^{N}\{\alpha_i^* t_i < \phi(x_i), \phi(x) >\} + b^*\}$$

f(x)=sign{∑i=1N{αi∗ti<ϕ(xi),ϕ(x)>}+b∗}

2.5 对参数 α, ξα,ξ 的讨论

基于上面的推导，可以很容易的看出，**软间隔最大化分类器** 就是在 **硬间隔最大化分类器** 的基础上，处理非线性可分的数据。而参数α, ξα,ξ起着决定样本点类别的作用，那么，这里就对这两个参数做一下讨论，看看这两个参数是如何判断非线性可分数据集中的样本点的类别的。

为了方便，这里直接用一张表格说明参数α, ξα,ξ 和样本点之间的关系，图种叉叉的部分，说明这种情况并不存在，下面将对这个表格做详细的说明：

		分类正确的点			分类错误的点
	ξ \ α	ξ=0	0<ξ<1	ξ=1	ξ>1
不是支持向量	α=0	非支持向量 对决策面没有贡献	×	×	×
是支持向量	0<α<C	支持向量 在SVM边界上	×	×	×
	α=C	×	支持向量 在SVM正确边界内	支持向量 在SVM超平面上	支持向量 分类错误点

在 2.1 软间隔最大化分类器的两个问题 中，就讨论过，非线性可分情况下，SVM中的样本点一共有五种情况，也就是上面表格中的所对应的五种情况。

> 从表格中可以粗略的观察到：
> - 当$\alpha_i = 0$αi=0 时，对应的样本点是非支持向量；当$\alpha_i > 0$αi>0 时，对应的样本点是支持向量
> - 当$\xi \le 1$ξ≤1 时，对应的样本点可以被正确的分类；当$\xi > 1$ξ>1 时，对应的样本点分类错误

现在基于参数 α、ξα、ξ 对上面五种情况再做更详细的讨论：

1. 当$\xi_i = 0, \alpha_i = 0$ξi=0,αi=0 时，此时的样本点 $x_i$xi 并不是支持向量，其被完全的分类正确，样本点满足：$t_i\{w^T\phi(x_i) + b\} > 1$ ti{wTϕ(xi)+b}>1。
变量关系：当$\alpha_i = 0$αi=0时，由KKT条件中的第三条：$C - \alpha_i - \mu_i = 0$C−αi−μi=0可以知道，$\mu_i = C$ μi=C；再由KKT条件第七条：$\mu_i^*\xi_i^* = 0$μi∗ξi∗=0 可以知道：$\xi_i = 0$ξi=0；最后根据KKT条件的第四条和第五条，可以很容易知道：$t_i\{w^T\phi(x_i) + b\} > 1$ ti{wTϕ(xi)+b}>1。

2.当$\xi_i = 0, 0 < \alpha_i < C$ξi=0,0<αi<C时,此时样本点刚好在SVM的间隔边界上，是支持向量，被正确分类，样本点满足：$t_i\{w^T\phi(x_i)+b\}=1$ ti{wTϕ(xi)+b}=1。
变量关系：当$0 < \alpha_i < C$ 0<αi<C 时，由KKT条件中的第三条：$C-\alpha_i-\mu_i=0$ C−αi−μi=0可以知道，$0 < \mu_i < C$ 0<μi<C;再由KKT条件第七条：$\mu_i^*\xi_i^*=0$μi∗ξi∗=0，可以知道：$\xi_i=0$ξi=0。最后根据KKT条件的第四条和第五条，可以很容易知道：$t_i\{w^T\phi(x_i)+b\}=1$ ti{wTϕ(xi)+b}=1。

3.当$0 < \xi_i < 1, \alpha_i = C$ 0<ξi<1,αi=C时，此时样本点在SVM的正确的间隔边界内，是支持向量，被正确分类，样本点满足：$0 < t_i\{w^T\phi(x_i)+b\} < 1$ 0<ti{wTϕ(xi)+b}<1。
变量关系：当$\alpha_i = C$ αi=C时，由KKT条件中的第三条：$C-\alpha_i-\mu_i=0$ C−αi−μi=0可以知道，$\mu_i=0$μi=0;再由KKT条件第六、七条：$\xi_i^* \geq 0$；$\mu_i^*\xi_i^*=0$ξi∗≥0；μi∗ξi∗=0，可以知道：$\xi_i > 0$ξi>0。根据KKT条件的第四条和第五条，可以很容易知道：$t_i\{w^T\phi(x_i)+b\}=1-\xi_i$ ti{wTϕ(xi)+b}=1−ξi 。那么，此时 $0 < \xi_i < 1$0<ξi<1，所以 $0 < t_i\{w^T\phi(x_i)+b\} < 1$ 0<ti{wTϕ(xi)+b}<1，即，其对应的样本点在分类正确的间隔中。

4.当$\xi_i = 1, \alpha_i = C$ ξi=1,αi=C 时，此时样本点在SVM决策面上，是支持向量，被正确分类，样本点满足：$t_i\{w^T\phi(x_i)+b\}=0$ ti{wTϕ(xi)+b}=0。
变量关系：这里的变量关系的分析和第三种情况一样，差别仅仅是 $\xi_i=1$ξi=1 ，由：$t_i\{w^T\phi(x_i)+b\}=1-\xi_i$ ti{wTϕ(xi)+b}=1−ξi，可以很容易知道：$t_i\{w^T\phi(x_i)+b\}=0$ ti{wTϕ(xi)+b}=0 ，这样，样本点x_i xi就刚好在SVM的决策面上了。

5.当$\xi_i > 1, \alpha_i = C$ ξi>1,αi=C时，是支持向量，被错误分类，样本点满足：$t_i\{w^T\phi(x_i)+b\} < 0$ ti{wTϕ(xi)+b}<0。
变量关系：这里的变量关系的分析和第三种情况一样，差别仅仅是 $\xi_i > 1$ξi>1 ，由：$t_i\{w^T\phi(x_i)+b\}=1-\xi_i$ ti{wTϕ(xi)+b}=1−ξi，可以很容易知道：$t_i\{w^T\phi(x_i)+b\} < 0$ ti{wTϕ(xi)+b}<0 ，这样，样本点x_i xi就在决策面错误的一边，被错误分类了。

由于变量αα是我们最终要求解的变量，这里再单独对变量αα做一个总结。首先，这里取$g(x)$g(x)如下，即SVM的决策面函数：

$$g(x)=\sum_{i=1}^{N}\{\alpha_i t_i < \phi(x_i), \phi(x) >\}+b$$

g(x)=∑i=1N{αiti<ϕ(xi),ϕ(x)>}+b

那么，根据前面的表格，可以很容易得到：

$$\begin{aligned} \alpha_i = 0 &\Leftrightarrow t_i g(x_i) \geq 1 \\ 0 < \alpha_i < C &\Leftrightarrow t_i g(x_i) = 1 \\ \alpha_i = C &\Leftrightarrow t_i g(x_i) \leq 1 \end{aligned}$$

αi=0⇔tig(xi)≥10<αi<C⇔tig(xi)=1αi=C⇔tig(xi)≤1

总的来说就是：

- 在SVM间隔区间外的点，对应的$\alpha_i=0$αi=0
- 在SVM间隔区间边缘上的点，对应的$0 < \alpha_i < C$ 0<αi<C
- 在SVM间隔区间内部的点，对应的$\alpha_i = C$ αi=C