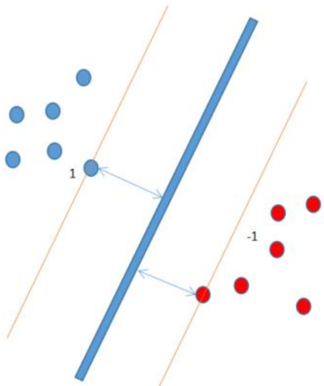
从优化建模角度看 SVM 的前世今生

一、SVM 基本原理

1.基本思想

SVM 的中心思想就是寻找一个平面将样本空间一分为二，完成二分类。对于二维空间的样本，我们分割数据需要一条线。对于三维空间，分割空间需要一个面，**进一步，对于n维空间，我们需要一个$n-1$维的超平面**。

观察下图，很明显数据可以分为两部分，理论上我们有无数种划分方法，那么哪种划分最好？SVM 认为划分的间隔越大越好，间隔也就是被超平面划分为两个部分之间最接近的距离间隔中间的那根蓝线就是 SVM 找到的用来划分的超平面。



同时我们可以发现，对于间隔，大多数样本不起作用，能够起作用的只有边缘的样本，这些样本确定了间隔，从而决定了超平面，我们把这些落在边缘的样本称为**支持向量**，这也是支持向量机（SVM，Support Vector Machine）的由来。

2.优化建模

我们首先来考虑最简单的情况，即线性可分，也就是说所有样本都可以被正确的划分。这样划分出来得到的间隔是实实在在的，所以我们把线性可分的情况下的间隔称为**硬间隔**。此时我们可以写出这个分隔平面的公式： $w^T x + b = 0$。

每个样本距离平面的距离即为$\gamma = \frac{|\omega^T x + b|}{||\omega||_2}$

对于样本点，在分隔平面上方的类别y_i为 1，下方的为-1，由此，我们可以得到$y(\omega^T x + b) > 0$。

我们来观察支持向量，也就是刚好在间隔边缘的点，它们到分割平面的距离刚好是间隔的一半。我们假设这个点的函数值是γ，我们把它表示出来可以得到：

$$y_i(\omega^T x_i + b) = \gamma$$

$$y_i\left(\frac{\omega}{\gamma} x_i + \frac{b}{\gamma}\right) = 1$$

由此我们可以将函数的值缩放为 1，便于后续计算，这样对于支持向量，就有$y(\omega^T x + b) = 1$，由此我们可以计算出间隔$\gamma = 2\frac{|\omega^T x + b|}{||\omega||_2} = \frac{2}{||\omega||_2}$,而 SVM 的目标即是让这个间隔尽量

大，所有这就是**一个带约束的优化问题**，如下所示：

$$\min_{\omega,b}\frac{1}{2}||\omega||^2$$
$$s.t. y_i(\omega^T x+b)\geq 1, i=1,2,3\dots,k$$

3.对偶问题求解

由于 SVM 的优化问题为凸优化问题且满足 Slater 条件，由此我们可以得到原始问题的解与其对偶问题的解相等，同时，我们可以**解其 KKT 条件**

首先是得到原问题的拉格朗日函数：

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

$$=\frac{1}{2}||\omega||^2+\sum_{i=1}^{m}(\alpha_i-\alpha_i y_i\omega^T x_i-\alpha_i y_i b)$$

再对ω和b进行求导:

$$\frac{\partial L}{\partial \omega}=\frac{1}{2}*2*\omega+0-\sum_{i=1}^{m}\alpha_i y_i x_i-0=0\rightarrow\omega=\sum_{i=1}^{m}\alpha_i y_i x_i$$

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

我们通过求导得出了ω和α之间的关系，也就是说只要我们确定了α也就确定了ω，另外我们可以发现上面的式子当中已经没有了b，说明b已经被我们消去了。我们把上面求导得到的取极值时的ω和b代入原式，消去ω和b，可以得到：

$$\min_{\omega,b}L(\omega,b,\alpha)=\frac{1}{2}\omega^T\omega+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_i y_i\omega^T x_i-\sum_{i=1}^{m}\alpha_i y_i b$$

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

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

观察以上式子，会发现x和y都是固定的值由样本确定，唯一的变量就只有α了。我们要求的是上式的极大值，唯一的变量是α，求出了α就可以推导得到对应的ω和b。

于是，SVM 问题便转化为了一下问题：

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

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

对于使用 SVM 分类我们便有了以下步骤:

1. 首先根据上面的最优化问题求出一系列的α

2.然后根据$\omega = \sum_{i=1}^{m} \alpha_i y_i x_i, b = \frac{1}{m}\sum_{j=1}^{m}\left(y_j - \sum_{i=1}^{m} \alpha_i y_i x_i^T x_j\right)$,求出参数$\omega$和$b$

3.当一个样本进来的时候，即根据得到的ω和b计算对应的y值，得到分类结果

二、软间隔

1.软间隔的定义

上述就数据实际可分的情况进行了讨论,但是实际上数据并不一定是百分百线性可分的,如果硬生生地找到一个分隔分开样本，反而可能陷入过拟合之中，反而不值得。**所以我们可以放松标准，允许部分样本出错**，为了解决这个问题，我们对原始的公式进行变形，**引入松弛变量**，这个松弛变量允许我们适当放松$y_i(\omega^T x_i + b) \geq 1$的条件为$y_i(\omega^T x_i + b) \geq 1 - \xi_i$。

对于一个松弛变量ξ_i，一个样本会有以下几种情况:

（1）$\xi = 0$，表示样本能够正确分类

（2）$0 < \xi < 1$，表示样本在分割平面和支持向量之间

（3）$\xi = 1$，表示样本在分割平面上

（4）$\xi \geq 1$，表示样本异常

因此，我们可以重写优化条件:

$$min \quad \frac{1}{2}\left||\omega|\right|^2 + C\sum_{i=1}^{m}\xi_i$$

$$s.t. y_i(\omega^T x_i + b) \geq 1 - \xi_i, i = 1,2,3 \dots, n$$
$$\xi_i \geq 0, i = 1,2,3 \dots, n$$

其中C是一个常数，可以看为一个惩罚系数，我们希望$\left||\omega|\right|^2$尽量小，同时也希望$\sum_{i=1}^{m}\xi_i$尽量小，这里C就是为了调节两者的平衡的，C越大表示对模型的分类越严格，越小表示对松弛的要求越低。

2.软间隔的对偶问题

和处理硬间隔的方法一样，我们通过**解优化问题的对偶问题**来解软间隔的优化问题。首先对于以上优化问题，我们得出拉格朗日函数为:

$$L(\omega, b, \xi, \alpha, \beta) = \frac{1}{2}\left||\omega|\right|^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\alpha_i\left(1 - \xi_i - y_i(\omega^T x_i + b)\right) - \sum_{i=1}^{m}\beta_i \xi_i$$

通过求导我们可以得到:

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

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

$$\frac{\partial L}{\partial \xi} = 0 \to \beta_i = C - \alpha_i$$

将以上三个式子代入拉格朗日函数，得到

$$L(\omega,b,\xi,\alpha,\beta)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j+C\sum_{i=1}^{m}\xi_i+\sum_{i=1}^{m}\alpha_i(1-\xi_i)-\sum_{i=1}^{m}(C-\alpha_i)\xi_i$$

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

由于$\beta_i \ge 0$，我们可以得到$0 \le \alpha_i \le C$，由此我们可以把对偶问题化为以下优化问题：

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

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

然后，我们回过来看 KKT 条件的其他部分：

$$1-\xi_i-y_i(\omega^T x_i+b)\le 0$$
$$-\xi_i \le 0$$
$$\alpha_i \ge 0$$
$$\beta_i=C-\alpha_i \ge 0$$
$$\alpha_i\left(1-\xi-y_i(\omega^T x_i+b)\right)=0$$
$$\beta_i\xi_i=0$$

我们考虑倒数第二个式子，无非两种情况，$\alpha_i=0$或者$1-\xi-y_i(\omega^T x_i+b)=0$，我们分情况讨论：

（1） 如果$\alpha_i=0$，那么$y_i(\omega^T x_i+b)-1\ge 0$，样本分类正确，不会对模型产生影响

（2） 如果$\alpha_i>0$，那么$y_i(\omega^T x_i+b)=1-\xi_i$，则样本是支持向量，由于$\beta_i=C-\alpha_i$和$\beta_i\xi_i=0$，我们又可以分为以下情况：

① $\alpha_i<C$，那么$\beta_i>0$，所以$\xi_i=0$，那么样本在边界上

② 如果$\alpha_i=C$，那么$\beta_i=0$，如果此时$\xi\le 1$，那么样本被正确分类，否则分类错误

由于化简后的式子中只剩下了α，后续只需要求出α即可求解出优化问题了，下一节中我们将介绍其数值求解方法，包括 SMO 算法等。

三、求解算法

1.SMO 算法

（1）基本思想：想要同时优化得到上述一系列α的值比较困难，所以 SMO（Sequential Minimal Optimization）算法提出了一种想法，**把α中的两个分量看为变量，其他的全部看成是常数。** 选择两个分量而不选择一个分量是为了保证满足$\sum_{i=1}^{m}\alpha_i y_i=0$的约束条件，即一个分量变化的时候，另一个分量也可以同时变化。此时，我们在上一节提到的优化问题即可化为以下问题，其中$K_{ij}=x_i^T x_j$

$$min_{\alpha 1,\alpha_2}\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2\alpha_1\alpha_2-(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{m}y_i\,\alpha_i K_{i1}+y_2\alpha_2\sum_{i=3}^{m}y_i\,\alpha_i K_{i,2}$$
$$+Constant$$

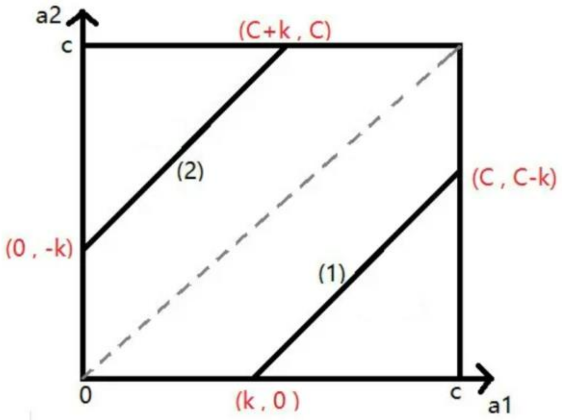
$$s.t.\alpha_1 y_1+\alpha_2 y_2=-\sum_{i=3}^{m} y_i\,\alpha_i$$

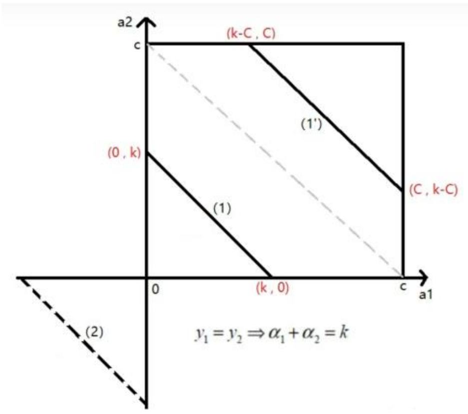
$$0\le\alpha_i\le C, i=1,2,3\ldots,m$$

由于$y_i=\pm1$，所以$y_i^2=1$，上面的$Constant$表示除了α_1,α_2以外的视为常数的项，我们假设$\alpha_1 y_1+\alpha_2 y_2=k$，其中$\alpha_1,\alpha_2\in[0,C]$，由于$y_i$只有 1 或-1 两种情况，于是我们可以分类讨论。

（2）分类讨论

①**首先是y_1,y_2不同号的时候**，也是两种情况，即$\alpha_1-\alpha_2=k$和$-\alpha_1+\alpha_2=k$，也就是$\alpha_2=\alpha_1\pm k$，由于对于α有约束条件的范围，于是这就变成了一个线性规划问题，如下方左图所示。针对图中的情况（1），有α_2的范围有$(0,C-\alpha_2+\alpha_1)$，情况（2）有$(\alpha_2-\alpha_1,C)$，我们这里使用迭代的方法来优化α_1,α_2，我们令上一轮的取值为α_{1o},α_{2o}，我们把刚才求得的综合一下，就可以得到α_2下一轮的下界L是$\max(0,\alpha_{2o}-\alpha_{1o})$，上界$H$是$\min(C+\alpha_{2o}-\alpha_{1o},C)$。





②同理，**我们讨论y_1,y_2同号的时候**，有上方右图的两种情况，与①类似，我们可以得到α_2下一轮的下界L是$\max(0,\alpha_{1o}+\alpha_{2o}-C)$，上界$H$是$\min(\alpha_{1o}+\alpha_{2o},C)$。

通过上面的讨论我们假设通过迭代之后得到的下一轮α_2是$\alpha_{2new,unc}$，这里的unc是未经过约束的意思。那么我们加上刚才的约束，可以得到：

$$\alpha_{2new}=\begin{cases} H, & \alpha_{2new,unc}>H \\ \alpha_{2new,unc}, & L\le\alpha_{2new,unc}\le H \\ L, & \alpha_{2new,unc}<L \end{cases}$$

这里$\alpha_{2new,unc}$是我们利用求导取到的极值α_2，但是由于约束的存在，这个极值不一定能取到。后续我们将以此为基础继续讨论。

（3）求解$\alpha_{2new,unc}$

我们已经得到了下一轮迭代之后的新的α_2的取值范围了，接下来，我们要做的就是不断迭代求解，求出使得目标函数最小的α_1,α_2，由于$\alpha_1+\alpha_2$已经确定，所以我们只需要求解其中一个即可。我们令$\alpha_1 y_1+\alpha_2 y_2=\xi$，于是我们有$\alpha_1=y_1(\xi-\alpha_2 y_2)$，把这个式子代入原始目标函数，将$\alpha_1$消去，我们得到：

$$W=\frac{1}{2}K_{11}(\xi-\alpha_2 y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+y_2K_{12}(\xi-\alpha_2 y_2)\alpha_2-(\xi-\alpha_2 y_2)y_1-\alpha_2+(\xi-\alpha_2 y_2)v_1 \\ +y_2\alpha_2 v_2$$

其中，我们令$v_i=\sum_{j=3}^{m}y_j\alpha_jK_{i,j}$,$E_i=f(x_i)-y_i=\sum_{j=1}^{m}\alpha_j\,y_jK_{i,j}+b-y_i$，$E_i$表示第$i$个样本真实值与预测值之差，然后，对原式求导求极值，得到：

$$\frac{\partial W}{\partial \alpha_2}=K_{11}\alpha_2+K_{22}\alpha_2-2K_{12}\alpha_2-K_{11}\xi y_2+K_{12}\xi y_2+y_1y_2-1-v_1y_2+y_2v_2=0$$

求解这个式子，我们最终得到：

$$\alpha_{2new,unc}=\alpha_{2o}+\frac{y_2(E_1-E_2)}{K_{11}+K_{22}-2K_{12}}$$

根据这个式子我们就可以求出下一轮迭代的值，求出之后，与上一节提到的上下限进行比较，即可得到满足约束下的取值，最后把α_2的代入约束条件，即可求出α_1，于是这样就是同时优化了一组参数，**SMO 算法就是不断重复上述过程，不停选择两个参数进行优化，直到达到迭代次数，或者不能再带来新的提升。**

2.其他算法

求解 SVM 问题同时也可以使用其他算法，比如**内点法（IPM）**，内点法使用对数阻挡函数将 SVM 的对偶问题由极大值问题转化为极小值问题。内点法由于需要计算N阶矩阵的逆，在使用牛顿迭代法时也需要计算$Hessian$矩阵的逆，所以内点法的内存开销较大且复杂度为$O(N^3)$，仅适用于少量学习样本的情形。

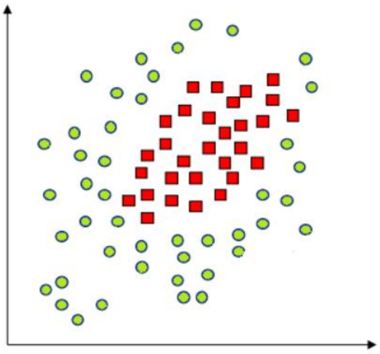
粒子群算法（PSO）也是一种能用于求解 SVM 优化问题的优化算法，其基本核心是利用群体中的个体对信息的共享从而使整个群体的运动在问题求解空间中产生从无序到有序的演化过程，从而获得问题的最优解。

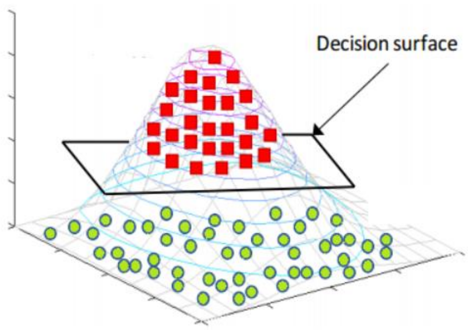
随机梯度算法（SGD）是机器学习问题中常见的优化算法，也可以用于求解 SVM 问题。

四、核函数

1.提出背景

以上几节我们已经将 SVM 模型的原理和求解方法完整推导出来了，那么为什么又要提出一个核函数呢？我们可以看下左图这个分类问题：





观察此图，我们就会发现，无论如何也不可能使用一条线将上面这个分类问题完成，因为上面分类问题的决策边界是一个椭圆。但是如果我们把上面的数据映射到更高的维度，也就是三维当中，我们就可以使用一个平面将样本分开了，如上右图所示，也就是说通过一个映射函数，**将样本从 n 维映射到 n+1 或者更高的维度，使得原本线性不可分的数据变成线性可分，**这样我们就解决了一些原本不能解决的问题。

2.核函数的定义

如果存在一个函数K，对应输入为$X \times X, X \in R$，对于任意$\forall x, z \in X$存在$\exists \phi: X \to Z$（将输入空间映射到更高维的空间Z），并且满足于条件$s.t.\ k(x,z) = < \phi(x), \phi(z) >$，则称$k(x,z)$为正定核函数。

3.常见的核函数

（1）高斯核函数（径向基核函数，RBF）

高斯核函数的定义为

$$k\left(x_i, x_j\right) = exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma^2\right) = exp\left(-\gamma \left\|x_i - x_j\right\|^2\right)$$

① 那么这个核函数将原始的空间升为多少维能，答案是**无穷维**，证明如下：

我们将高斯核函数展开，得到：

$$k(x1, x2) = exp(-(x_1^2 + x_2^2 - 2x_1x_2)/2\sigma^2)$$
$$= exp(-x_1^2/2\sigma^2) \times exp(-x_2^2/2\sigma^2) \times exp(-x_1x_2/\sigma^2)$$

我们重点关注其中的第三项，根据e^x的泰勒展开公式，我们将第三项展开得到：

$$1 - \frac{x_1}{\sigma}\frac{x_2}{\sigma} + \cdots \pm \frac{(x_1/\sigma)^n (x_2/\sigma)^n}{\sqrt{n!}\sqrt{n!}}$$

这一项看一看成两个向量的相乘的形式：

$$\left[1, \frac{x_1}{\sigma}, \ldots\right]\left[1, -\frac{x_2}{\sigma}, \ldots\right]$$

由此可以看出此核函数将原始空间上升为无穷维。

② 参数带宽σ的影响

假设我们将两个点x_1, x_2映射到无穷空间中，得到$\phi(x_1), \phi(x_2)$ 在高维空间中两个点之间的欧式距离为

$$d = \left|\left|\phi(x_1) - \phi(x_2)\right|\right|^2 = \phi(x_1)^2 + \phi(x_2)^2 - 2\phi(x_1)\phi(x_2)$$

根据核函数的定义可知，

$$\phi(x_1)^2 = k(x_1, x_1) = exp(0)$$

这样距离便化简为

$$d = 2 - 2 \times exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma^2\right)$$

由此可以看出：

当σ 为无穷时，$d \to 0$，在高维空间中对样本点的区分程度下降，容易造成欠拟合。

当σ 为 0 时， $d \to 2$，在高维空间中对样本点的区分度最高，容易造成过拟合。

（2）其他核函数

一些常见的其他核函数还有：

① 线性核函数，其实就是没有核函数。我们表示出来就是$k\left(x_i, x_j\right) = x_i^T x_j$

② 多项式核函数，它等价于一个多项式变换：$k\left(x_i, x_j\right) = \left(\gamma x_i^T x_j + b\right)^d$,这里的$\gamma$,$b$和$d$都是我们设置的参数

③ sigmoid 核，它的公式是：$k\left(x_i, x_j\right) = \tanh\left(\gamma x_i^T x_j + b\right)$

3.使用方法

现在我们已经知道了核函数是什么，将其运用到 SVM 中只需要将原来样本通过ϕ映射到高维，使用核函数计算即可，改进后的优化问题如下所示：

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

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

$$0\le\alpha_i\le C, i=1,2,3\dots,m$$

五、支持向量回归（SVR）

将 SVM 由分类推广至回归问题可以得到支持向量回归（SVR），分类 SVM（即 SVC）中的超平面决策边界即是 SVR 的回归模型：$f(X)=w^T X+b$，若样本点与回归模型足够接近，即落入回归模型的间隔边界内，则该样本不计算损失，对应的损失函数被称为ε-不敏感损失函数（ε-insensitive loss）：$L(z)=max(0,|z|-\epsilon)$，其中ϵ是决定间隔边界宽度的超参数。类比 SVM，SVR 是如下形式的二次凸优化问题：

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

$$s.t.|y_i-f(X)|\le\epsilon$$

使用松弛变量ξ,ξ^*表示ε-不敏感损失函数的分段取值后可得：

$$min_{w,b}\ \frac{1}{2}||w||^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)$$

$$s.t.y_i-f(X)\le\epsilon+\xi_i$$

$$f(X)-y_i\le\epsilon+\xi_i^*$$

$$\xi\ge 0,\xi^*\ge 0$$

同样，可以通过引入拉格朗日乘子可得到其拉格朗日函数：

$$L(w,b,\xi,\xi^*,\alpha,\alpha^*,\mu,\mu^*)=\frac{1}{2}||w||^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)-\sum_{i=1}^{N}\mu_{i\xi_i}-\sum_{i=1}^{N}\mu_i^*\xi_i^*$$

$$+\sum_{i=1}^{N}\alpha_i[f(X_i)-y_i-\epsilon-\xi_i]+\sum_{i=1}^{N}\alpha_i^*[f(X_i)-y_i-\epsilon-\xi_i^*]$$

通过对偶问题和求解 KKT 条件可以得到 SVR 的形式为：

$$f(X)=\sum_{i=1}^{m}(\alpha_i^*-\alpha_i)X_i^T X+b$$

SVR 也可以使用核函数得到非线性回归的结果。

六、SVM 的其他应用

1. 多分类 SVM：上述主要讨论的是 SVM 进行二分类的情况，SVM 也可以用于多分类，主

要方法有成对分类方法和一对余分类法。成对分类法的主要思想是每两个类之间都构造一个二分类 SVM，以获得对于多个分类问题的决策边界。一对余分类的主要思想是对于每一类，在这一类和其他类之间构造一个二分类 SVM，以获得多个分类问题的决策边界。

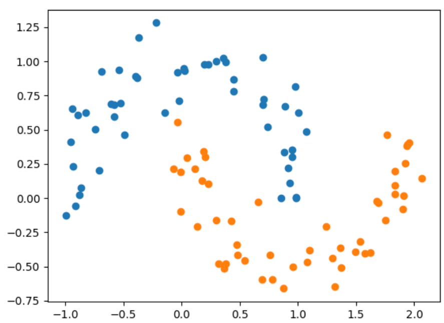
2.支持向量聚类：支持向量聚类是一类非参数的聚类算法，是 SVM 在聚类问题中的推广。具体地，支持向量聚类首先使用核函数，通常是 RBF，将样本映射至高维空间，随后使用 SVDD（Support Vector Domain Description）算法得到一个闭合超曲面作为高维空间中样本点富集区域的刻画。最后，支持向量聚类将该曲面映射回原特征空间 ，得到一系列闭合等值线，每个等值线内部的样本会被赋予一个类别。

3.SVM 也常常和其他机器学习算法（比如，LDA、PCA 等）共同使用以解决一些较为复杂的问题，比如用于人脸识别等。

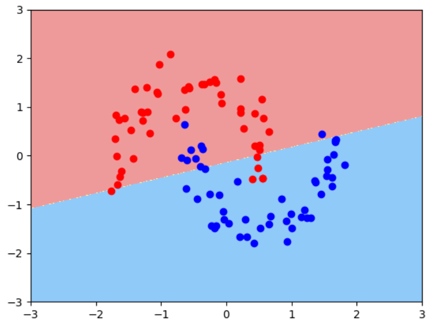
七、实验（使用 sklearn）

1.线性 SVM 与非线性 SVM

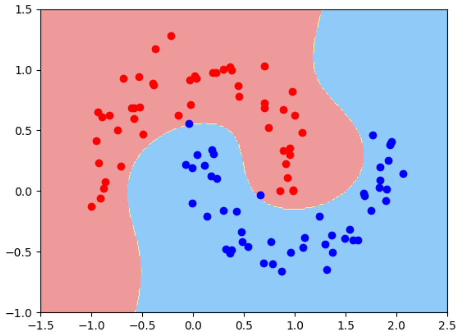
我们随机生成一个月亮型的二维二类数据集，而后分别使用线性 SVM、多项式核函数 SVM 和 RBF 核函数 SVM 进行二分类，画出决策边界，结果如下：



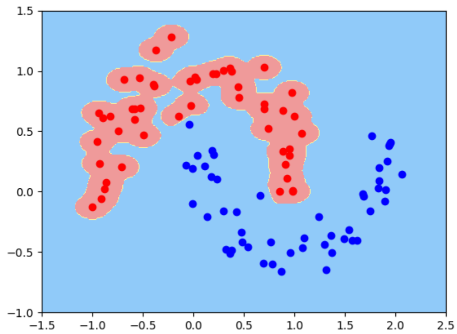
（1）原始数据



（2）线性 SVM 分类结果



（3）多项式核函数 SVM 分类结果



（4）RBF 核函数分类结果

从中可以看出核函数对于无法线性可分的数据的分类性能的优化。

2.使用 SVM 进行多分类

我们使用鸢尾花数据集进行多分类，使用数据集的 70%进行训练，30%进行验证，使用 RBF 核函数，最终得到在训练集上的准确率 96.2%，测试集上的准确率 97.8%，部分决策边界示意图如下：