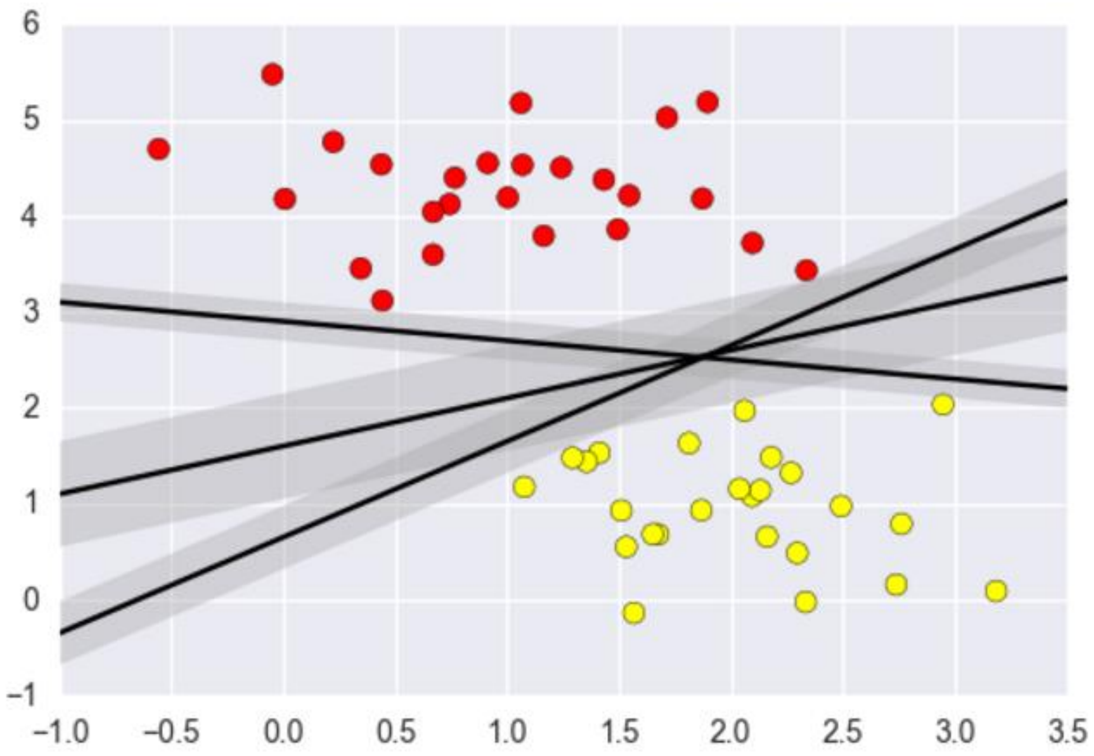
支持向量机

- Support Vector Machine

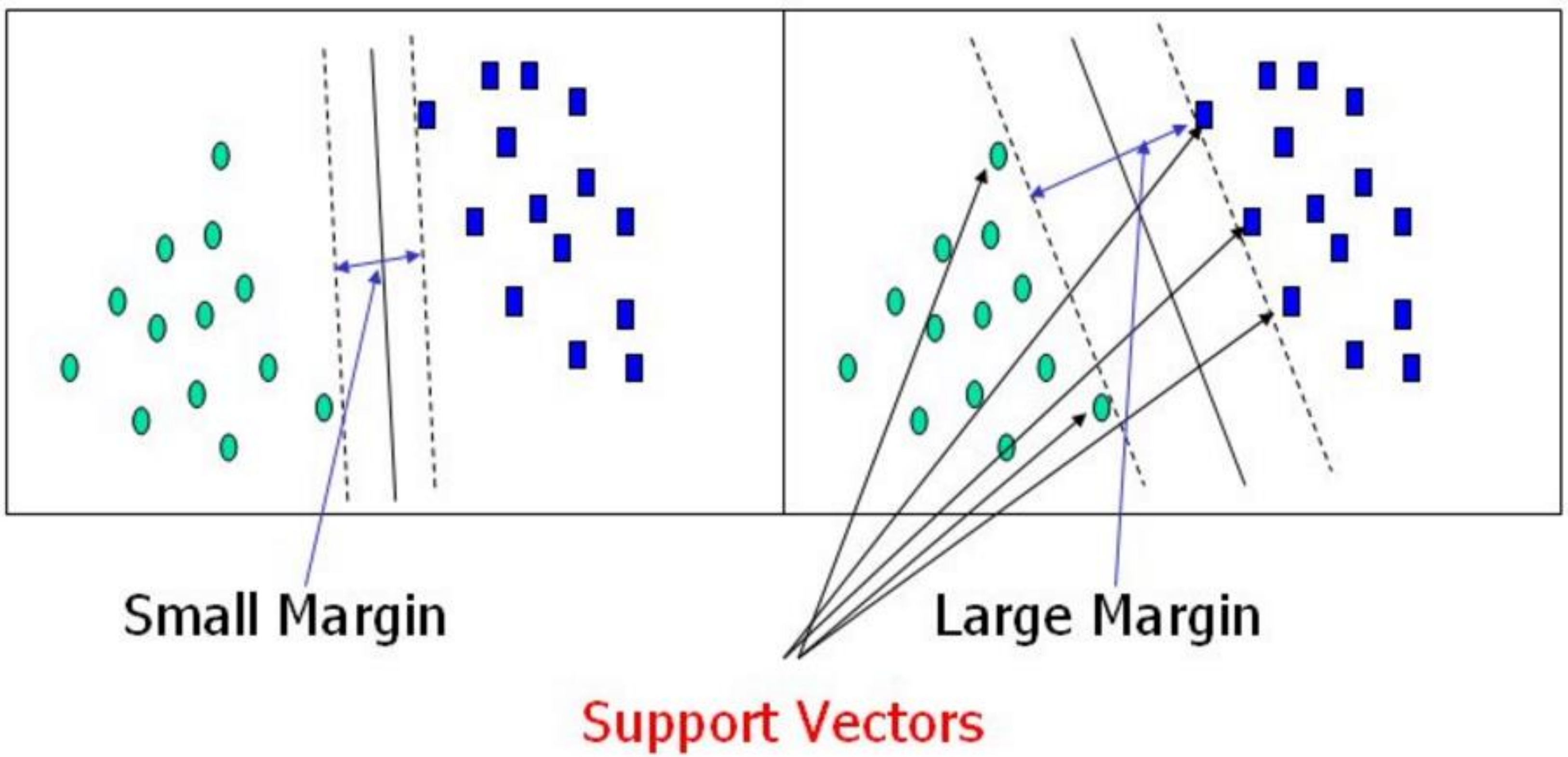
- 要解决的问题：什么样的决策边界才是最好的呢？
- 特征数据本身如果就很难分，怎么办呢？
- 计算复杂度怎么样？能实际应用吗？
- 目标：基于上述问题对SVM进行推导



支持向量机

- Support Vector Machine

  - 决策边界：选出来离雷区最远的（雷区就是边界上的点，要Large Margin）



支持向量机

距离的计算

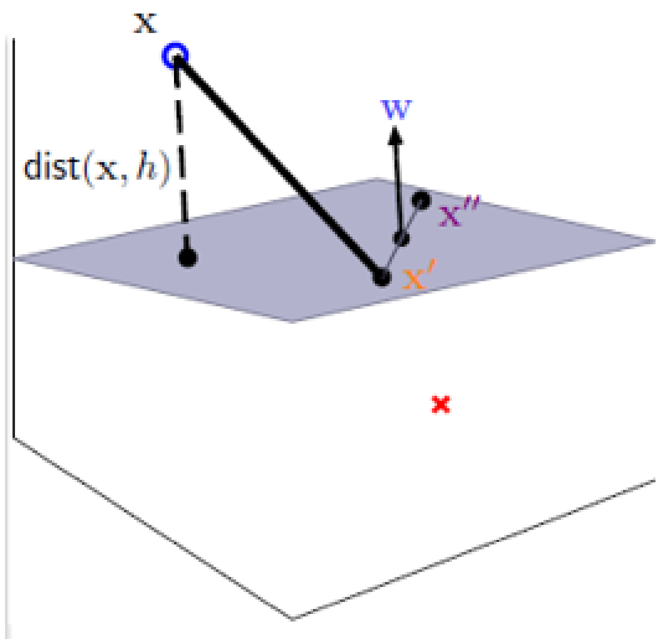
consider $\mathbf{x}'$, $\mathbf{x}''$ on hyperplane

① $\mathbf{w}^T\mathbf{x}' = -b$, $\mathbf{w}^T\mathbf{x}'' = -b$

② $\mathbf{w} \perp$ hyperplane:

$$\left(\mathbf{w}^T \underbrace{(\mathbf{x}'' - \mathbf{x}')}_{\text{vector on hyperplane}}\right) = 0$$

③ distance = project $(\mathbf{x} - \mathbf{x}')$ to $\perp$ hyperplane



$$\text{distance}(\mathbf{x}, b, \mathbf{w}) = \left|\frac{\mathbf{w}^T}{\|\mathbf{w}\|}(\mathbf{x} - \mathbf{x}')\right| \overset{①}{=} \frac{1}{\|\mathbf{w}\|}|\mathbf{w}^T\mathbf{x} + b|$$

支持向量机

- 数据标签定义

  - 数据集：(X1,Y1)(X2,Y2)... (Xn,Yn)

  - Y为样本的类别： 当X为正例时候 Y = +1 当X为负例时候 Y = -1

  - 决策方程：$y(x) = w^T \Phi(x) + b$ （其中 $\Phi(x)$ 是对数据做了变换，后面继续说）

$$=> \begin{matrix} y(x_i) > 0 \Leftrightarrow y_i = +1 \\ y(x_i) < 0 \Leftrightarrow y_i = -1 \end{matrix} \quad => \quad y_i \cdot y(x_i) > 0$$

支持向量机

优化的目标

- 通俗解释：找到一个条线（w和b），使得离该线最近的点（雷区）能够最远

- 将点到直线的距离化简得：$\dfrac{y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right)}{\|w\|}$

（由于 $y_i \cdot y(x_i) > 0$ 所以将绝对值展开原始依旧成立）

支持向量机

- 目标函数

放缩变换：对于决策方程（w,b）可以通过放缩使得其结果值|Y|>= 1

=> $y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right) \geq 1$ （之前我们认为恒大于0，现在严格了些）

优化目标：$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_i \left[y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right) \right] \right\}$

由于 $y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right) \geq 1$，只需要考虑 $\arg\max_{w,b} \frac{1}{\|w\|}$ （目标函数搞定！）

支持向量机

目标函数

- 当前目标：$max_{w,b}\ \frac{1}{||W||}$，约束条件：$y_i\left(w^T\cdot\Phi(x_i)+b\right)\geq 1$
- 常规套路：将求解极大值问题转换成极小值问题=>$min_{w,b}\ \frac{1}{2}w^2$
- 如何求解：应用拉格朗日乘子法求解

支持向量机

拉格朗日乘子法

带约束的优化问题： $\min_x \ f_0(x)$

$$subject\ to \quad f_i(x) \le 0,\ i = 1, \ldots m$$
$$h_i(x) = 0, i = 1, \ldots q$$

原式转换： $\min \ L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{q} v_i h_i(x)$

我们的式子： $L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(w^T \cdot \Phi(x_i) + b \right) - 1 \right)$

（约束条件不要忘： $y_i \left(w^T \cdot \Phi(x_i) + b \right) \ge 1$ ）

支持向量机

SVM求解

- 分别对w和b求偏导,分别得到两个条件（由于对偶性质）

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

- 对w求偏导： $\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i \Phi(x_n)$

对b求偏导： $\frac{\partial L}{\partial b} = 0 \Rightarrow 0 = \sum_{i=1}^{n} \alpha_i y_i$

支持向量机

SVM求解

带入原始：$L(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i(y_i(w^T\Phi(x_i)+b)-1)$

其中 $w=\sum_{i=1}^{n}\alpha_i y_i \Phi(x_n)$　　　$0=\sum_{i=1}^{n}\alpha_i y_i$

$$=\frac{1}{2}w^T w-w^T\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i)-b\sum_{i=1}^{n}\alpha_i y_i+\sum_{i=1}^{n}\alpha_i$$

$$=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}(\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i))^T\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i)$$

→ 完成了第一步求解 $\min_{w,b} L(w,b,\alpha)$

$$=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1,j=1}^{n}\alpha_i\alpha_j y_i y_j\Phi^T(x_i)\Phi(x_j)$$

支持向量机

SVM求解

继续对α求极大值：$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \left(\Phi(x_i)\cdot\Phi(x_j)\right)$

条件：$\sum_{i=1}^{n}\alpha_i y_i = 0$

$\alpha_i \geq 0$

极大值转换成求极小值：$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \left(\Phi(x_i)\cdot\Phi(x_j)\right) - \sum_{i=1}^{n}\alpha_i$

条件：$\sum_{i=1}^{n}\alpha_i y_i = 0$

$\alpha_i \geq 0$

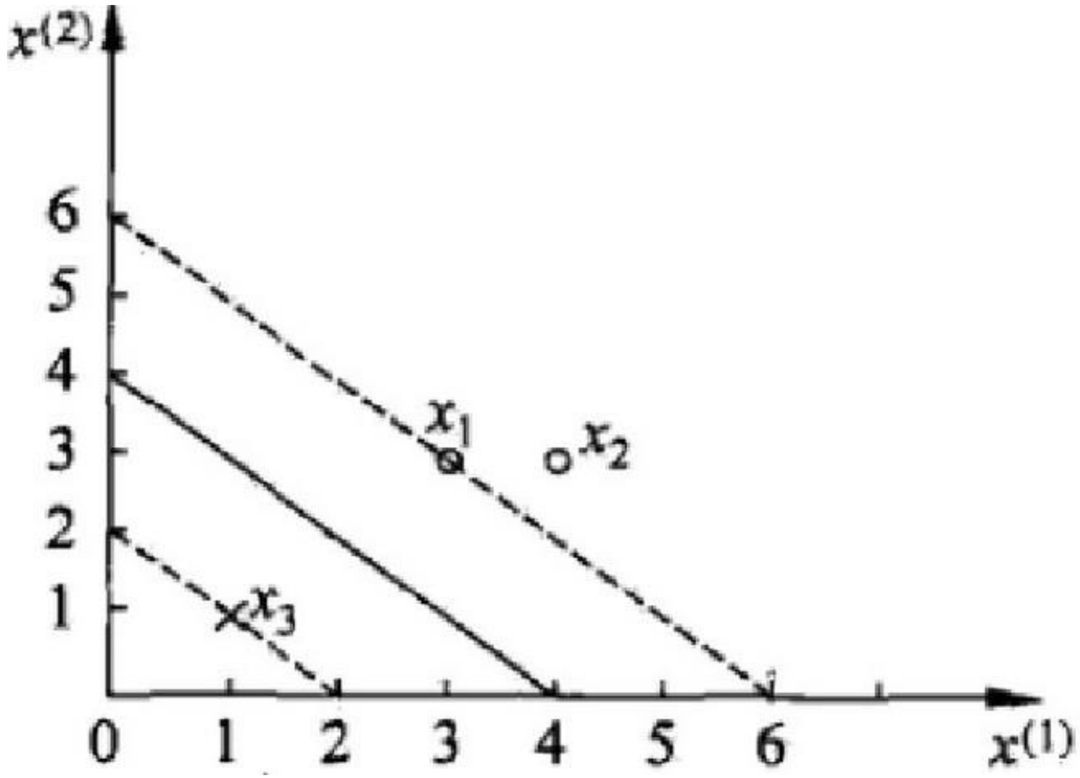
支持向量机

- SVM求解实例

数据：3个点，其中正例 X1(3,3) ，X2(4,3) ，负例X3(1,1)

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

约束条件：
$$\alpha_1 + \alpha_2 - \alpha_3 = 0$$
$$\alpha_i \geq 0, \quad i = 1,2,3$$



支持向量机

SVM求解实例

原式：$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i \cdot x_j\right)-\sum_{i=1}^{n}\alpha_i$ ，将数据代入

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

由于：$\alpha_1+\alpha_2=\alpha_3$ 化简可得：$4\alpha_1^2+\frac{13}{2}\alpha_2^2+10\alpha_1\alpha_2-2\alpha_1-2\alpha_2$

支持向量机

SVM求解实例

分别对α1和α2求偏导，偏导等于0可得：α_1 =1.5
α_2 =-1

（并不满足约束条件 $\alpha_i \geq 0, \quad i=1,2,3$ ，所以解应在边界上）

α_1 =0
α_2 =-2/13 ➡ 带入原式=-0.153 （不满足约束）

α_1 =0.25
α_2 =0 ➡ 带入原式=-0.25　（满足啦！）

最小值在(0.25,0,0.25)处取得

支持向量机

SVM求解实例

将α结果带入求解 $w=\sum_{i=1}^{n}\alpha_i y_i \Phi(x_n)$

$$w = \frac{1}{4} * 1 * (3,3) + \frac{1}{4} * (-1) * (1,1) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

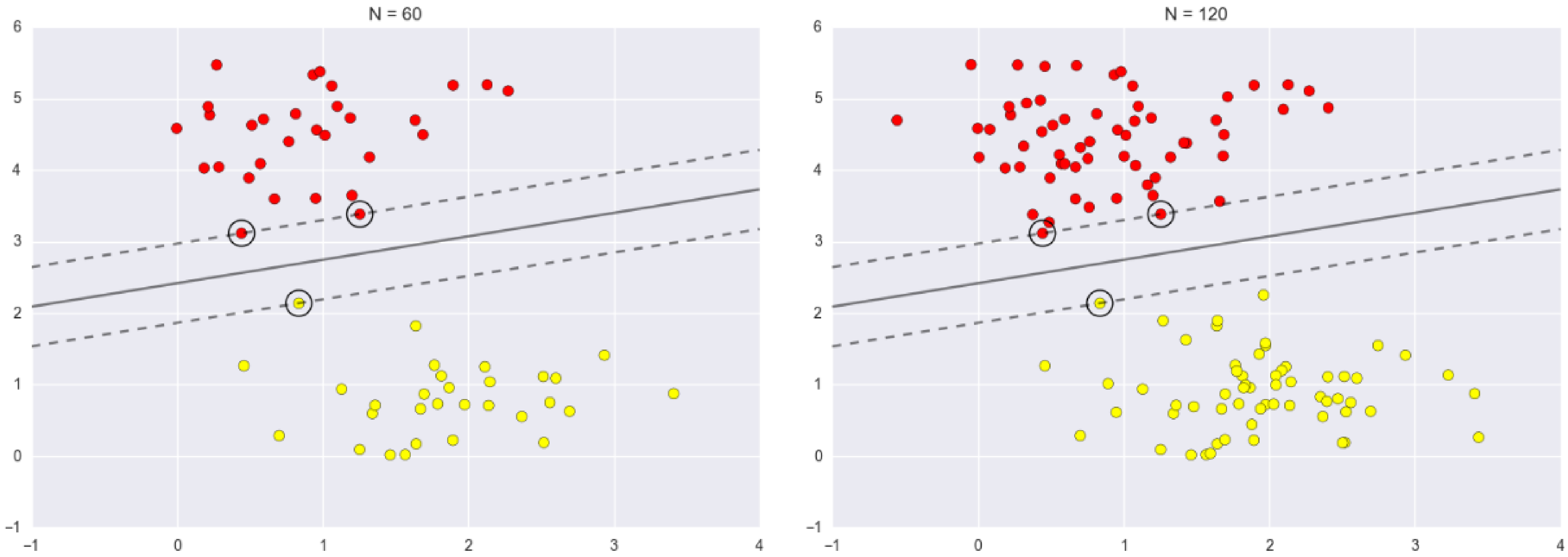
$$b = y_i - \sum_{i=1}^{n} a_i\, y_i\, (x_i\, x_j) = 1 - \left(\frac{1}{4} * 1 * 18 + \frac{1}{4} * (-1) * 6\right) = -2$$

平面方程为：$0.5x_1 + 0.5x_2 - 2 = 0$

支持向量机

SVM求解实例

支持向量：真正发挥作用的数据点，α值不为0的点



支持向量机

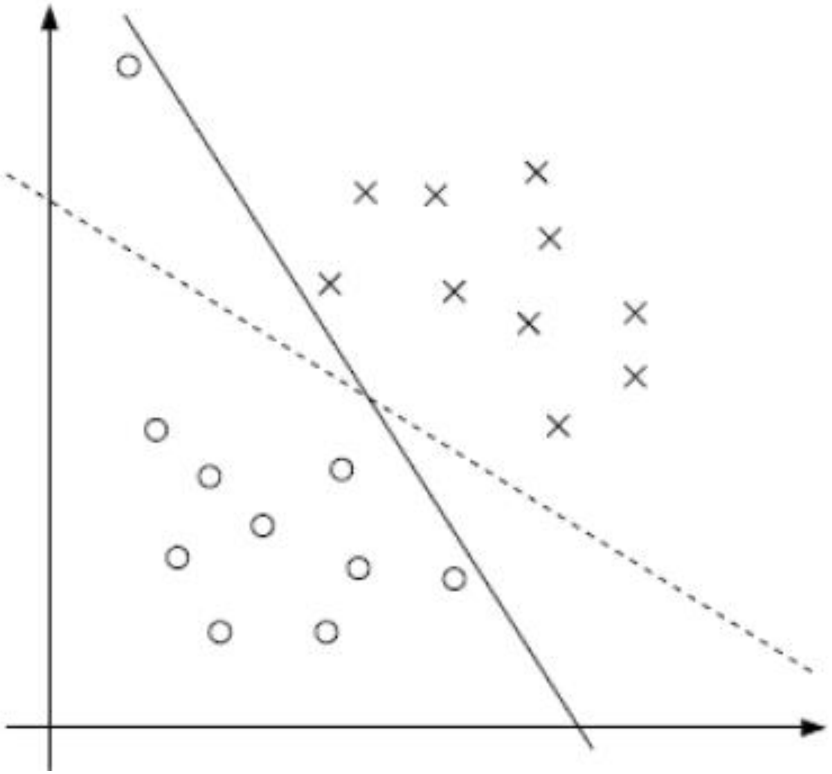
- soft-margin

  - 软间隔：有时候数据中有一些噪音点，如果考虑它们咱们的线就不太好了

  - 之前的方法要求要把两类点完全分得开，这个要求有点过于严格了，我们来放松一点！

  - 为了解决该问题，引入松弛因子

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



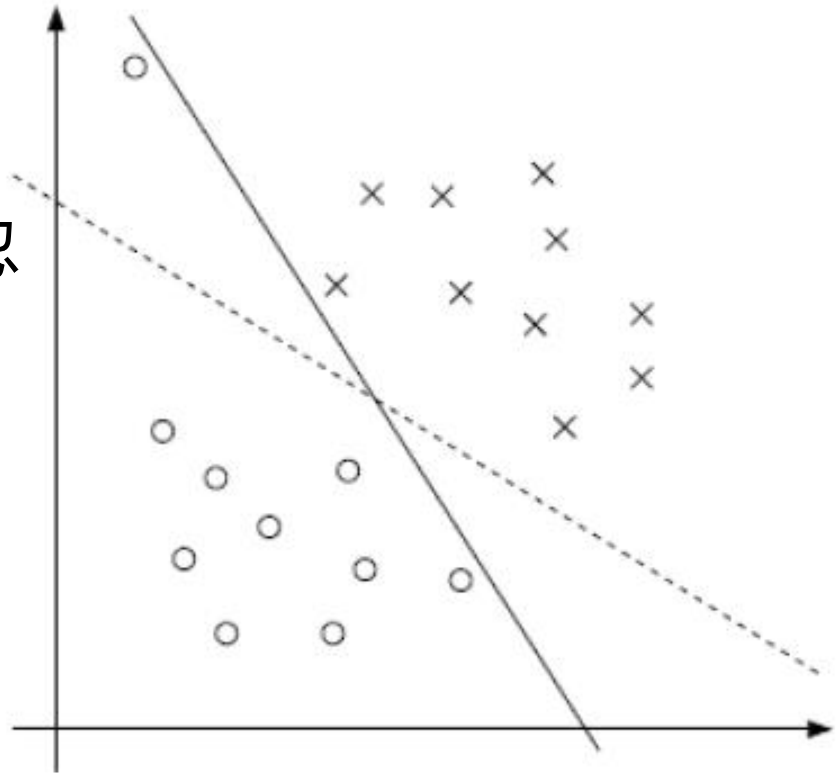
支持向量机

- soft-margin

  - 新的目标函数： $\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$

  - 当C趋近于很大时：意味着分类严格不能有错误
    当C趋近于很小时：意味着可以有更大的错误容忍

  - C是我们需要指定的一个参数！



支持向量机

- soft-margin

拉格朗日乘子法：

$$L(w,b,\xi,\alpha,\mu) \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(w\cdot x_i + b\right) - 1 + \xi_i\right) - \sum_{i=1}^{n}\mu_i\xi_i$$

$$w = \sum_{i=1}^{n}\alpha_i y_i \phi(x_n)$$

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

约束：$0 = \sum_{i=1}^{n}\alpha_i y_i$

同样的解法：$\sum_{i=1}^{n}\alpha_i y_i = 0$

$$C - \alpha_i - \mu_i = 0$$

$$0 \le \alpha_i \le C$$

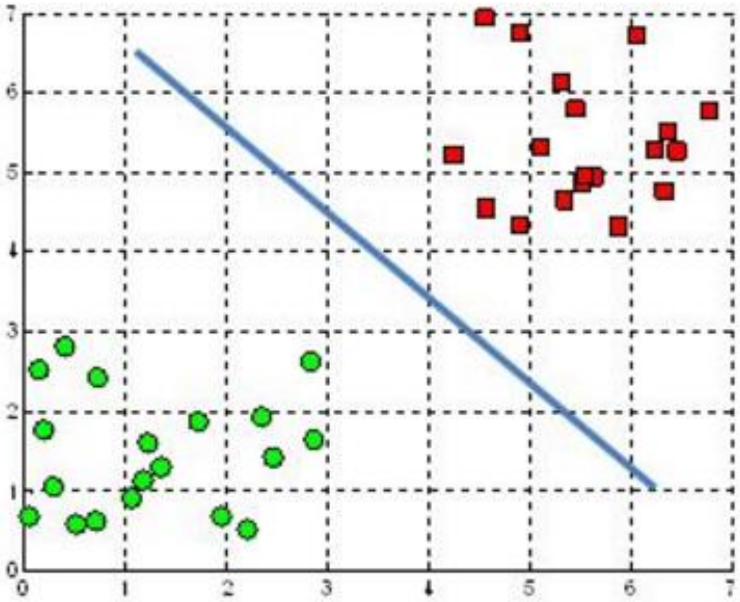
$$\alpha_i \ge 0 \quad \mu_i \ge 0$$

支持向量机

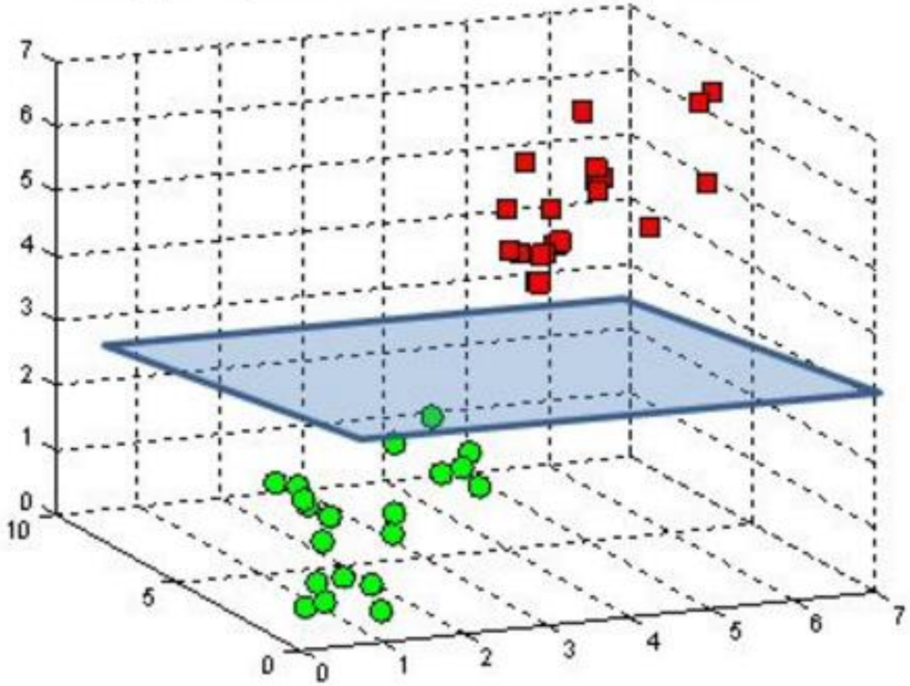
- 低维不可分问题

核变换：既然低维的时候不可分，那我给它映射到高维呢？

A hyperplane in $\mathbb{R}^2$ is a line



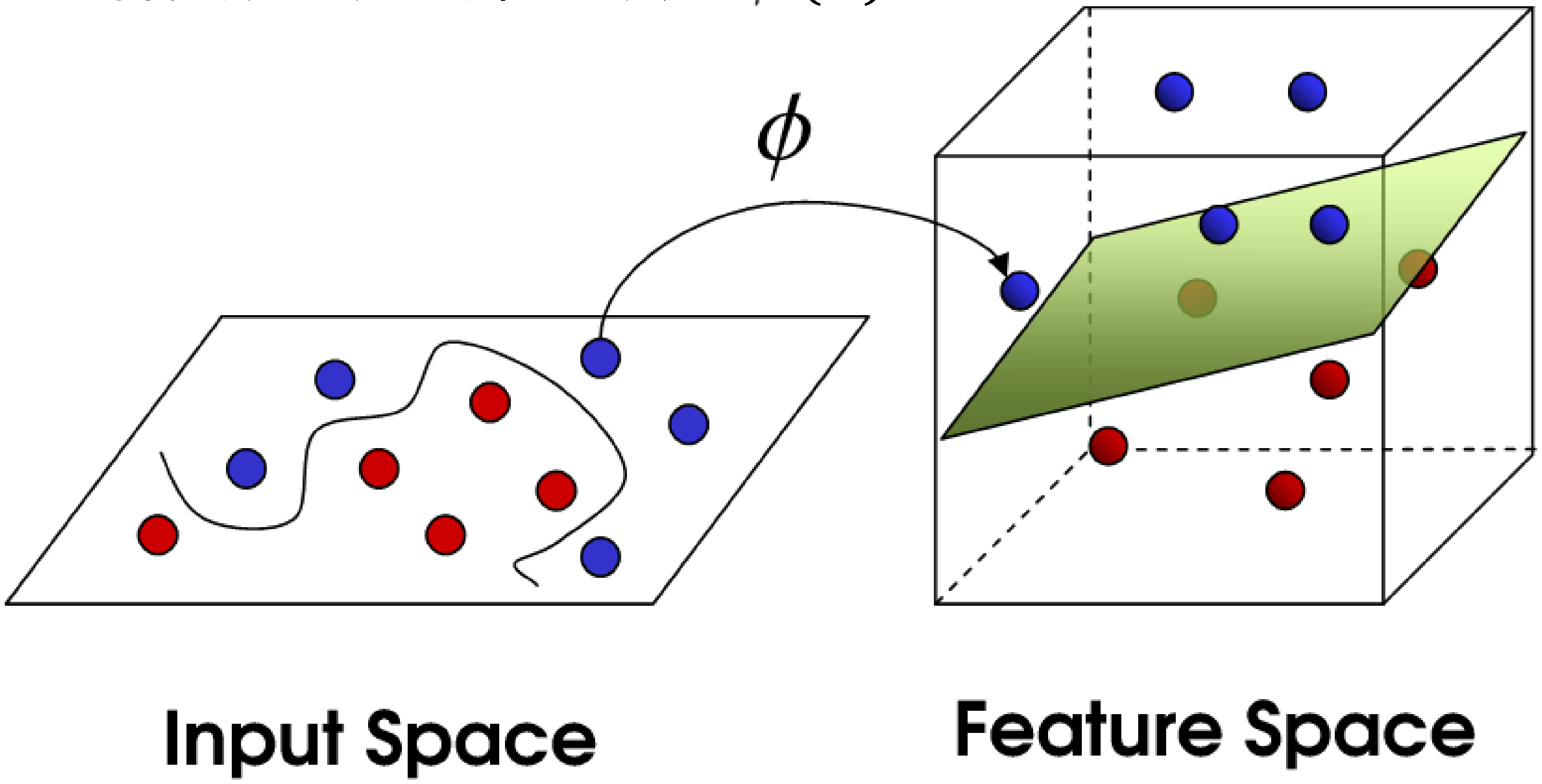
A hyperplane in $\mathbb{R}^3$ is a plane



支持向量机

- 低维不可分问题

目标：找到一种变换的方法，也就是 $\phi(X)$



支持向量机

还是先从一个小例子来阐述问题。假设我们有俩个数据，x = (x1, x2, x3); y = (y1, y2, y3)，此时在3D空间已经不能对其经行线性划分了，那么我们通过一个函数将数据映射到更高维的空间，比如9维的话，那么 f(x) = (x1x1, x1x2, x1x3, x2x1, x2x2, x2x3, x3x1, x3x2, x3x3)，由于需要计算内积，所以在新的数据在9维空间，需要计算<f(x),f(y)>的内积，需要花费O（n^2）。

在具体点，令x = (1, 2, 3); y = (4, 5, 6)，那么f(x) = (1, 2, 3, 2, 4, 6, 3, 6, 9)，f(y) = (16, 20, 24, 20, 25, 36, 24, 30, 36)，

此时<f(x), f(y)> = 16 + 40 + 72 + 40 + 100+ 180 + 72 + 180 + 324 = 1024

似乎还能计算，但是如果将维数扩大到一个非常大数时候，计算起来可就不是一丁点问题了。

但是发现， K(x, y) = (<x, y>)^2

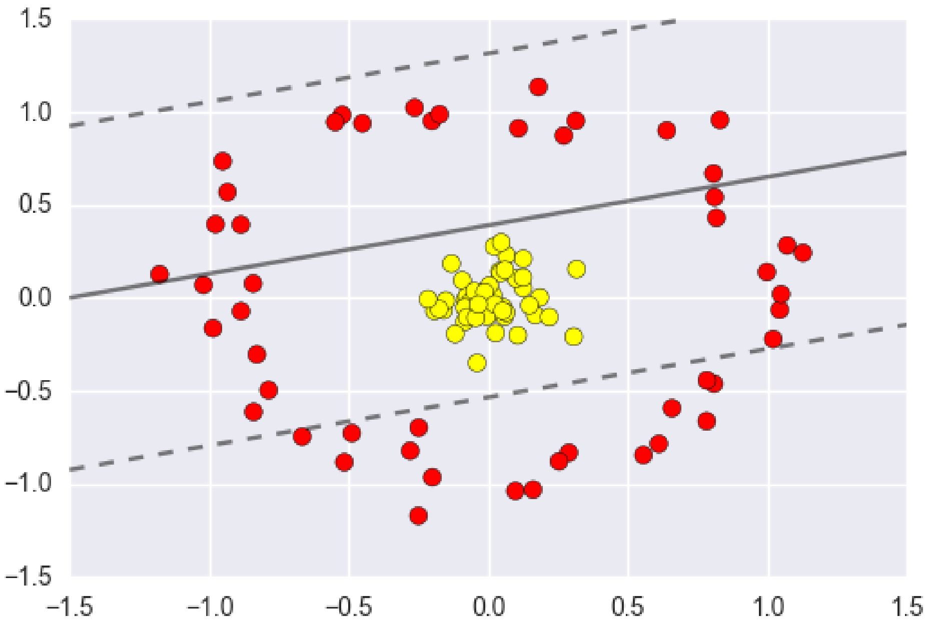
K(x,y)=(4 + 10 + 18) ^2 = 32^2 = 1024

俩者相等，K(x, y) = (<x, y>)^2=<f(x), f(y)>，但是 K(x, y) 计算起来却比<f(x), f(y)>简单的多，也就是说只要用K(x, y)来计算，，效果和<f(x), f(y)>是一样的，但是计算效率却大幅度提高了，如：K(x, y)是O（n），而<f(x),f(y)>是O（n^2）.所以使用核函数的好处就是，可以在一个低维空间去完成高维度（或者无限维度）样本内积的计算，比如K(x,y)=(4 + 10 + 18) ^2 的3D空间对比<f(x), f(y)> = 16 + 40 + 72 + 40 + 100+ 180 + 72 + 180 + 324 的9D空间。

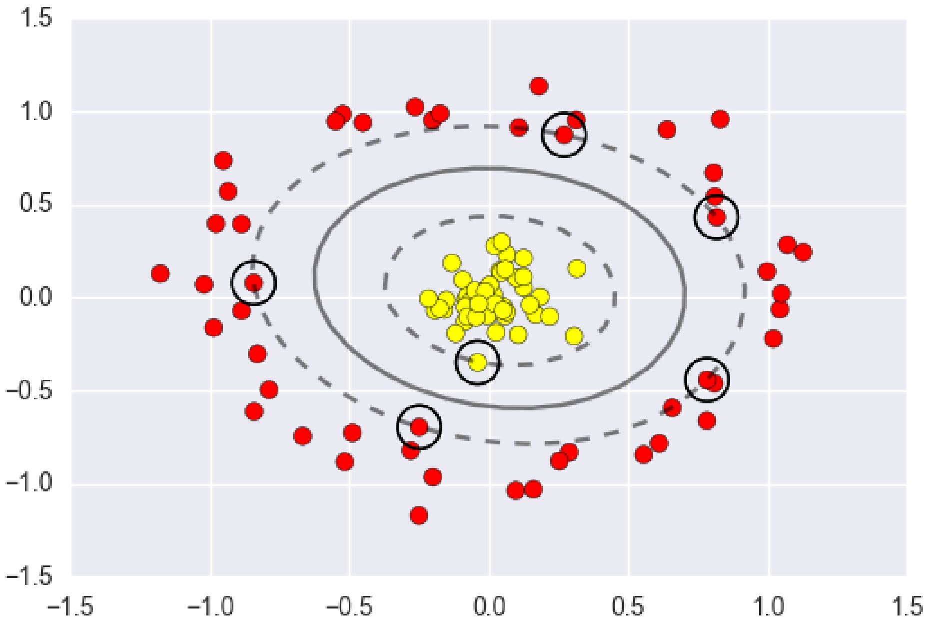
支持向量机

- Support Vector Machine

高斯核函数：$K(X,Y)=\exp\left\{-\frac{\|X-Y\|^2}{2\sigma^2}\right\}$



线性核函数



高斯和函数