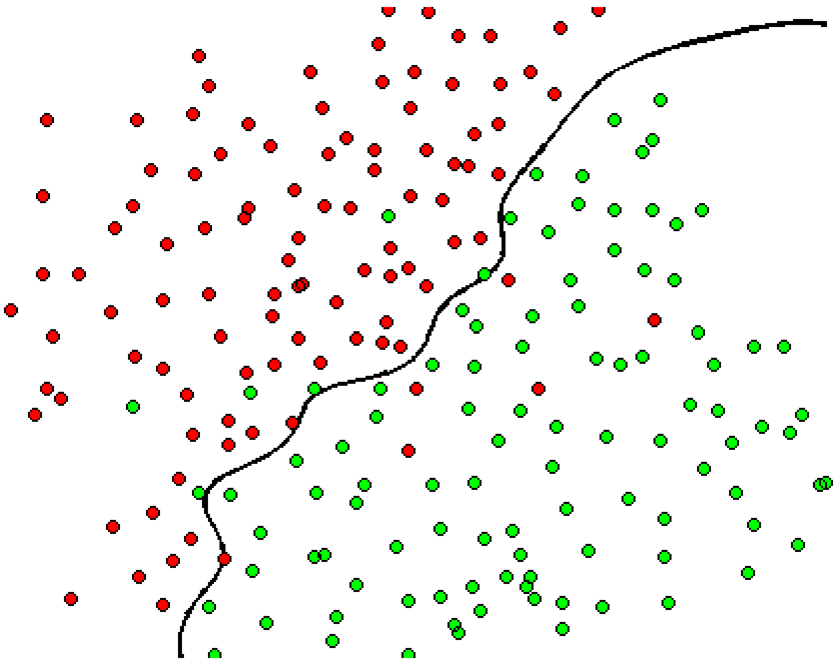
法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

关注 **小象学院**



分类模型(2)

--Robin

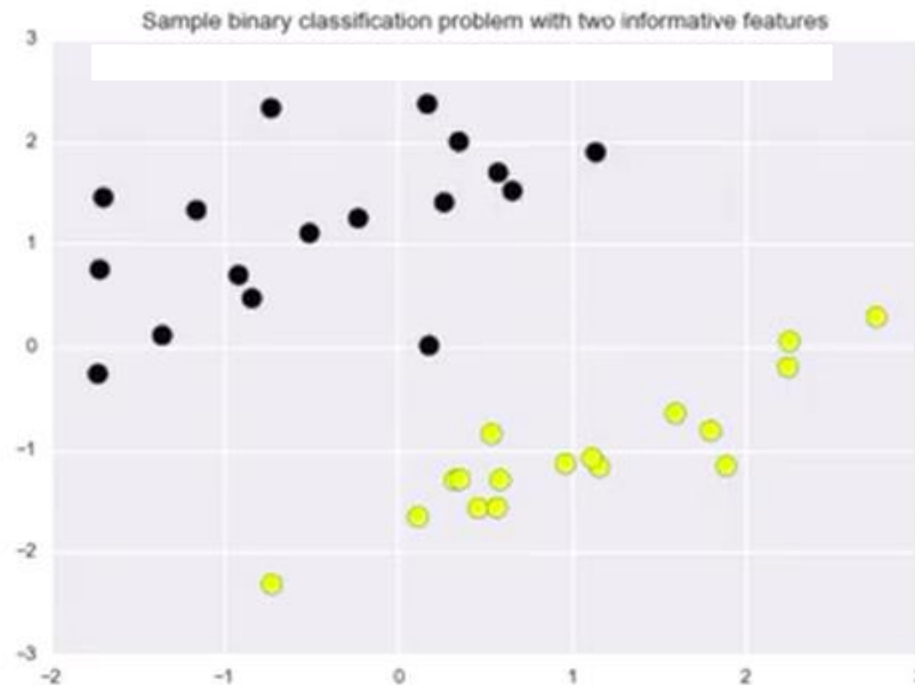
目录

- 逻辑回归
- 正则化
- 支持向量机

SVM

$$f(x, w, b) = sign(w \circ x + b)$$

$$= sign\left(\sum w[i]x[i] + \text{b}\right)$$



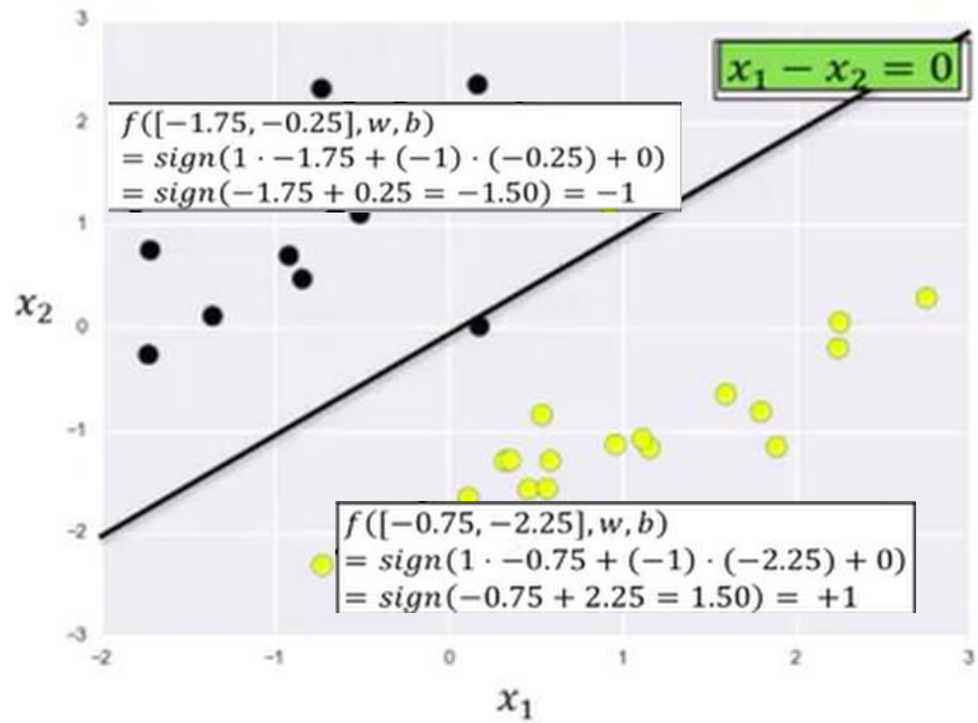
目标：找到一个可以分割这两个类别的边界

SVM

- 例子，假设这个边界已知

$$f(x, w, b) = sign(w \circ x + b)$$

$$x_1 - x_2 = 0$$
$$w = [1, -1]$$
$$b = 0$$



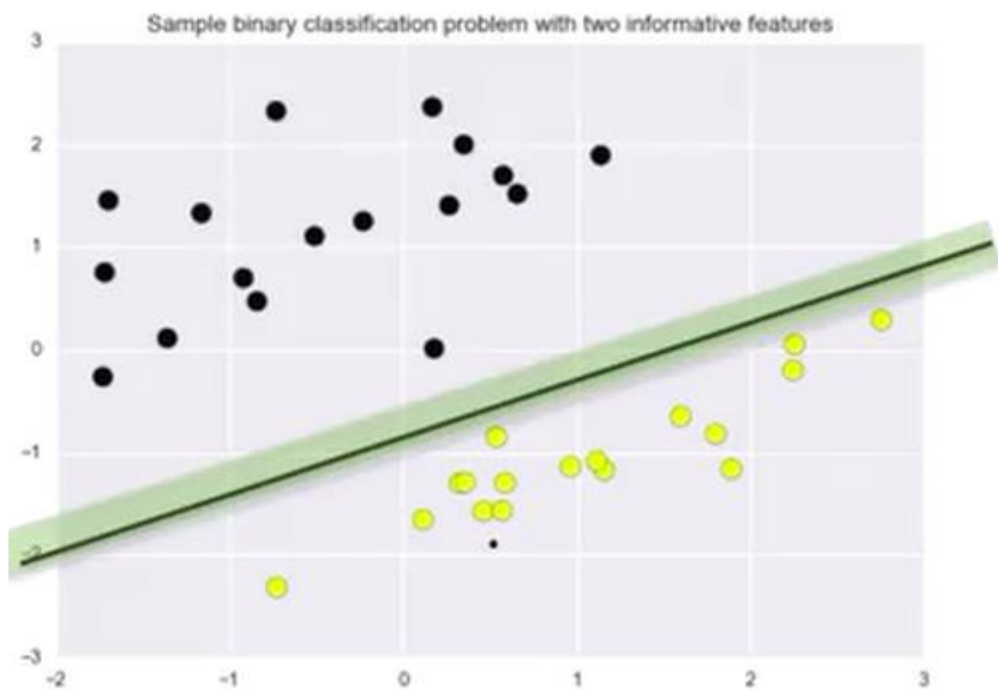
SVM

- 间隔（ Margin ）

$$f(x,w,b) = sign(w \circ x + b)$$

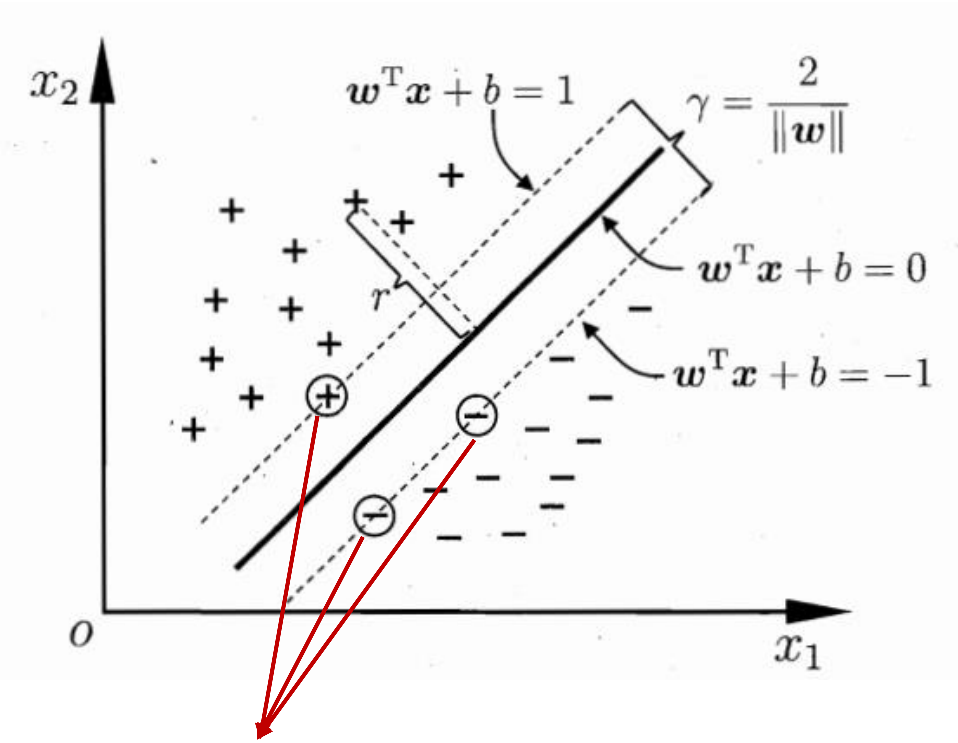
分类器间隔:

- 分类边界可以扩展到样本点的最大宽度。
- 例子中分类器的间隔就是**绿色区域**



SVM

- 间隔（ Margin ）



支持向量

- 带不等式约束的优化问题

$$\max_{\boldsymbol{w},b} \frac{2}{||\boldsymbol{w}||}$$
$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1, \quad i=1,2,\dots,m.$$

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1, \quad i=1,2,\dots,m.$$

SVM

- 拉格朗日乘子法

$$\min_{\boldsymbol{w},b} \ \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t.} \ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geqslant 1, \quad i = 1, 2, \ldots, m.$$

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)\right)$$

KKT条件

$$\begin{cases} \alpha_i \geqslant 0\ ; \\ y_i f(\boldsymbol{x}_i) - 1 \geqslant 0\ ; \\ \alpha_i \left(y_i f(\boldsymbol{x}_i) - 1\right) = 0 \end{cases}$$

其中 $\boldsymbol{\alpha} = (\alpha_1; \alpha_2; \ldots; \alpha_m)$，令*L*对*w*和*b*的偏导为零可得：

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \qquad 0 = \sum_{i=1}^{m} \alpha_i y_i$$

将*w*带入*L*中，可得

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

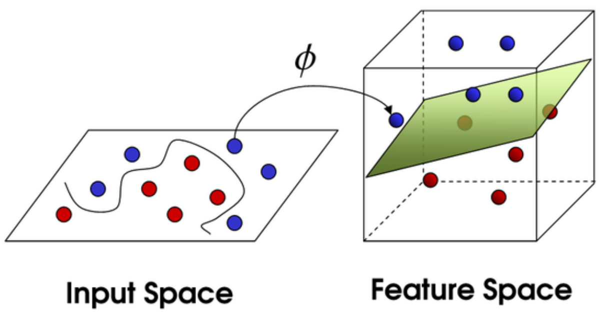
$$\text{s.t.} \ \sum_{i=1}^{m} \alpha_i y_i = 0\ ,$$
$$\alpha_i \geqslant 0\ , \quad i = 1, 2, \ldots, m\ .$$

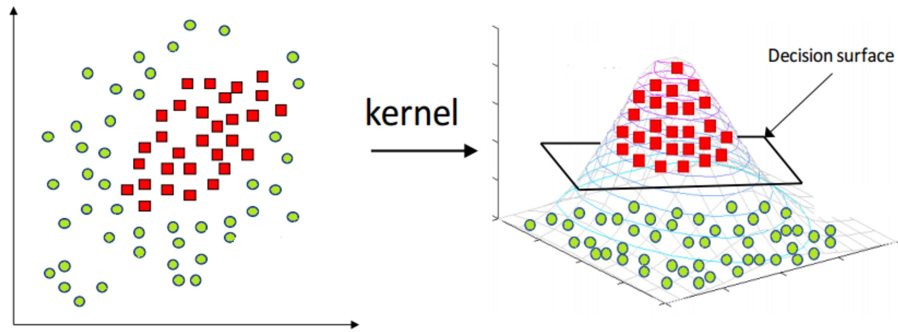
二次规划问题：可以使用SMO算法进行求解

SVM

核函数

- 如果数据线性不可分，怎么办？





- 将样本从原始空间映射到一个更高维的特征空间，从而使得样本在特征空间内线性可分

原始空间

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$$

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

特征空间

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}) + b$$

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

SVM

特征空间

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}) + b$$

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

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j y_i y_j \boxed{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)}$$

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

如果特征空间维数很高，直接计算会非常困难

核函数

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

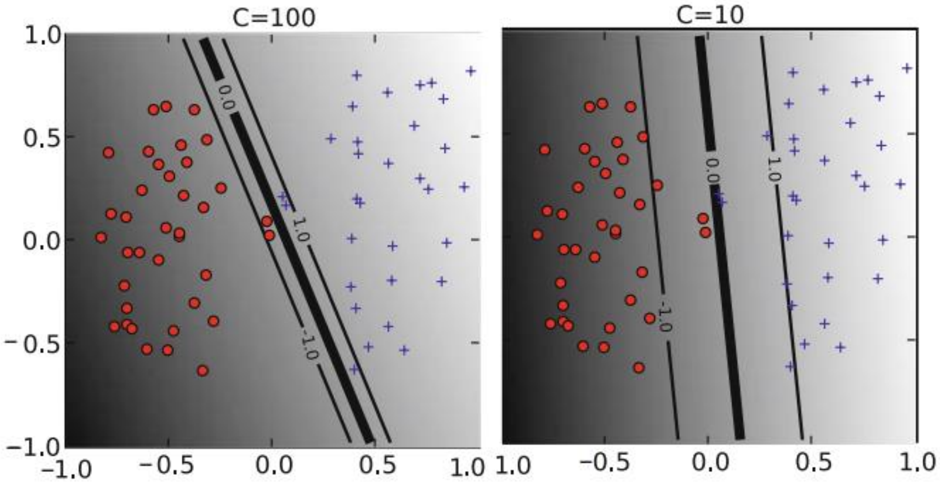
常用的核函数

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)^d$	$d \geqslant 1$ 为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert\boldsymbol{x}_i - \boldsymbol{x}_j\rVert^2}{2\sigma^2}\right)$	$\sigma > 0$ 为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert\boldsymbol{x}_i - \boldsymbol{x}_j\rVert}{\sigma}\right)$	$\sigma > 0$
Sigmoid 核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j + \theta)$	tanh 为双曲正切函数, $\beta > 0$, $\theta < 0$

SVM

$$\boxed{\min_f \sum_{i=1}^{n} V(f(x_i), y_i)} + \lambda \boxed{R(f)}$$

损失函数 正则项



C=1/λ

- 正则项中的C值决定了正则化的强度
- C值越大，正则化越弱 -> narrow margin
- C值越小，正则化越强 ->large margin

```
sklearn.svm.SVC
```

联系我们

小象学院：互联网新技术在线教育领航者

– 微信公众号：**小象学院**