svm汇报

蓝精灵儿

——感谢群里的各位老师

SVM

- 支持向量机(supportvectormachines．SVM)
- 二类分类模型.它的基本模型是定义在特征空间上的间隔最大的线性分类器，间隔最大使它有别于感知机；
- 支持向量机还包括核技巧，这使它成为实质上的非线性分类器.
- 支持向量机的学习策略就是间隔最大化，可形式化为一个求解凸二次规划(convex quadratic programming)的问题，也等价于正则化的合页损失函数的最小化问题.支持向量机的学习算法是求解凸二次规划的最优化算法.

SVM分类

- 支持向量机(supportvectormachines. SVM)
- 线性可分支持向量机(linear support vector machine in linearly separable case).
  - 硬间隔最大化(hard margin maximization);
- 线性支持向量机(linear supportvector machine)
  - 训练数据近似线性可分时，通过软间隔最大化(soft margin maximization);
- 非线性支持向量机(non-linear support vector machine)
  - 当训练数据线性不可分时，通过使用核技巧(kernel trick)及软间隔最大化

SVM分类

- 支持向量机(supportvectormachines．SVM)
- 当输入空间为欧氏空间或离散集合、特征空间为希尔伯特空间时，核函数(kernel function)表示将输入从输入空间映射到特征空间得到的特征向量之间的内积；
- 通过使用核函数可以学习非线性支持向量机，等价于隐式地在高维的特征空间中学习线性支持向量机，这样的方法称为核技巧；
- 核方法(kernel method)是比支持向量机更为一般的机器学习方法.

线性可分支持向量机

- 二分类问题：
- 输入空间：欧式空间或离散集合
- 特征空间：欧式空间或希尔伯特空间
- 线性可分支持向量机、线性支持向量机：假设这两个空间的元素一一对应，并将输入空间中的输入映射为特征空间中的特征向量；
- 非线性支持向量机：利用一个从输入空间到特征空间的非线性映射将输入映射为特征向量；
- 支持向量机的学习是在特征空间进行的.

线性可分支持向量机

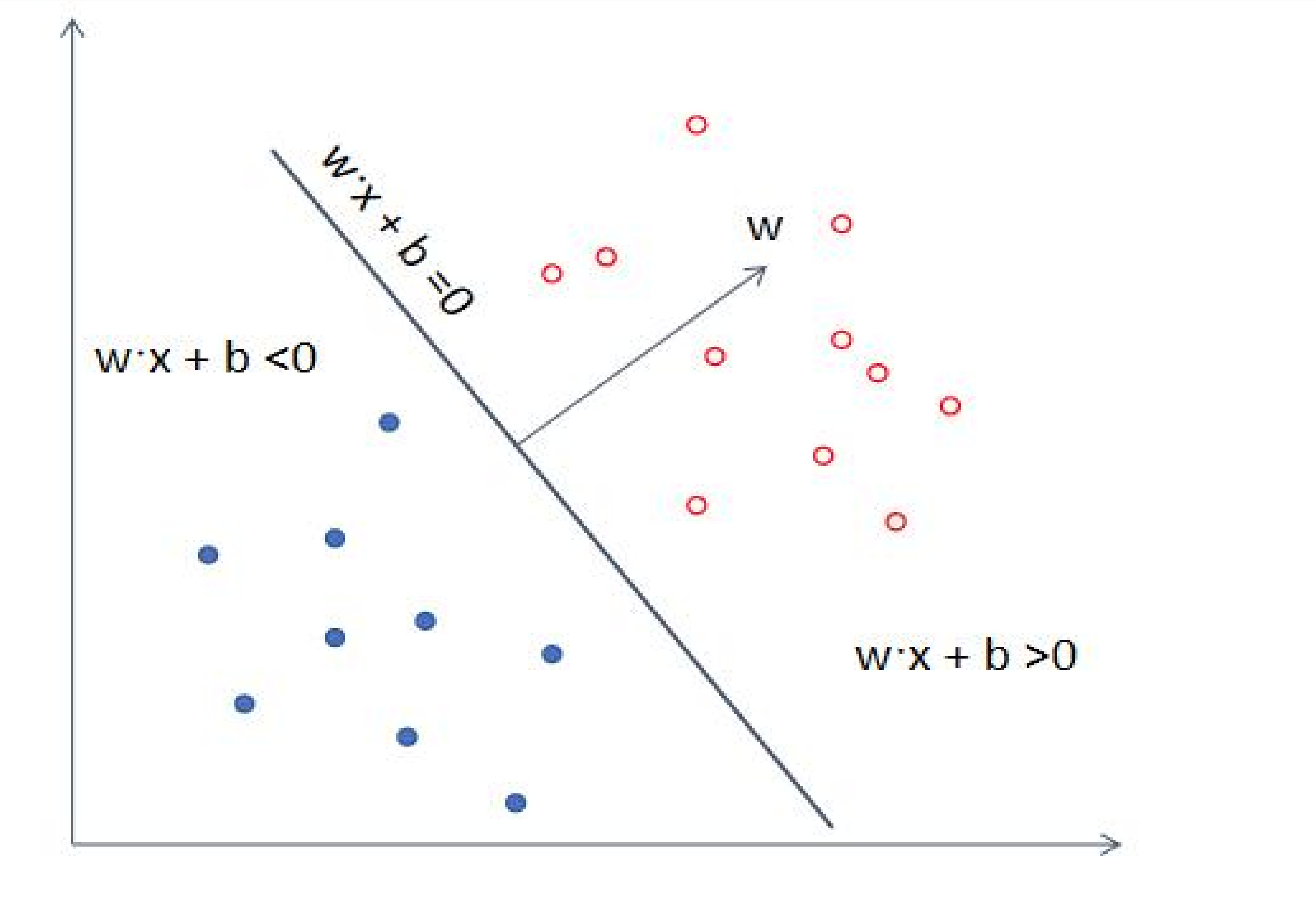
- 假设特征空间上的训练数据集：

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\} \qquad x_i\in\mathcal{X}=\mathbf{R}^n,\ y_i\in\mathcal{Y}=\{+1,-1\},\ i=1,2,\cdots,N$$

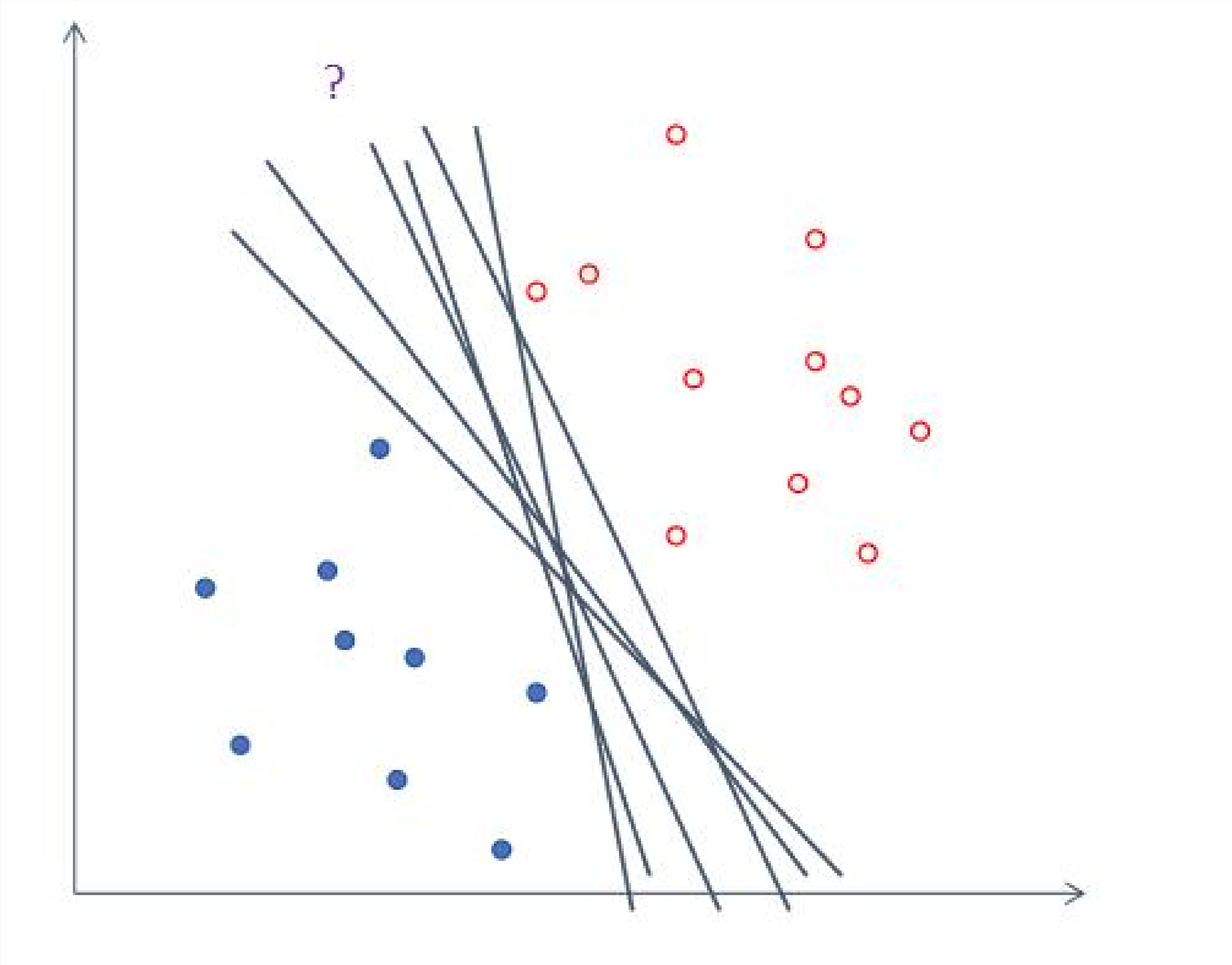
- 正例和负例
- 学习的目标：找到分类超平面，
- 线性可分支持向量机：给定线性可分训练数据集，通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为
- 决策函数：

$$w^*\cdot x+b^*=0 \qquad f(x)=\mathrm{sign}(w^*\cdot x+b^*)$$

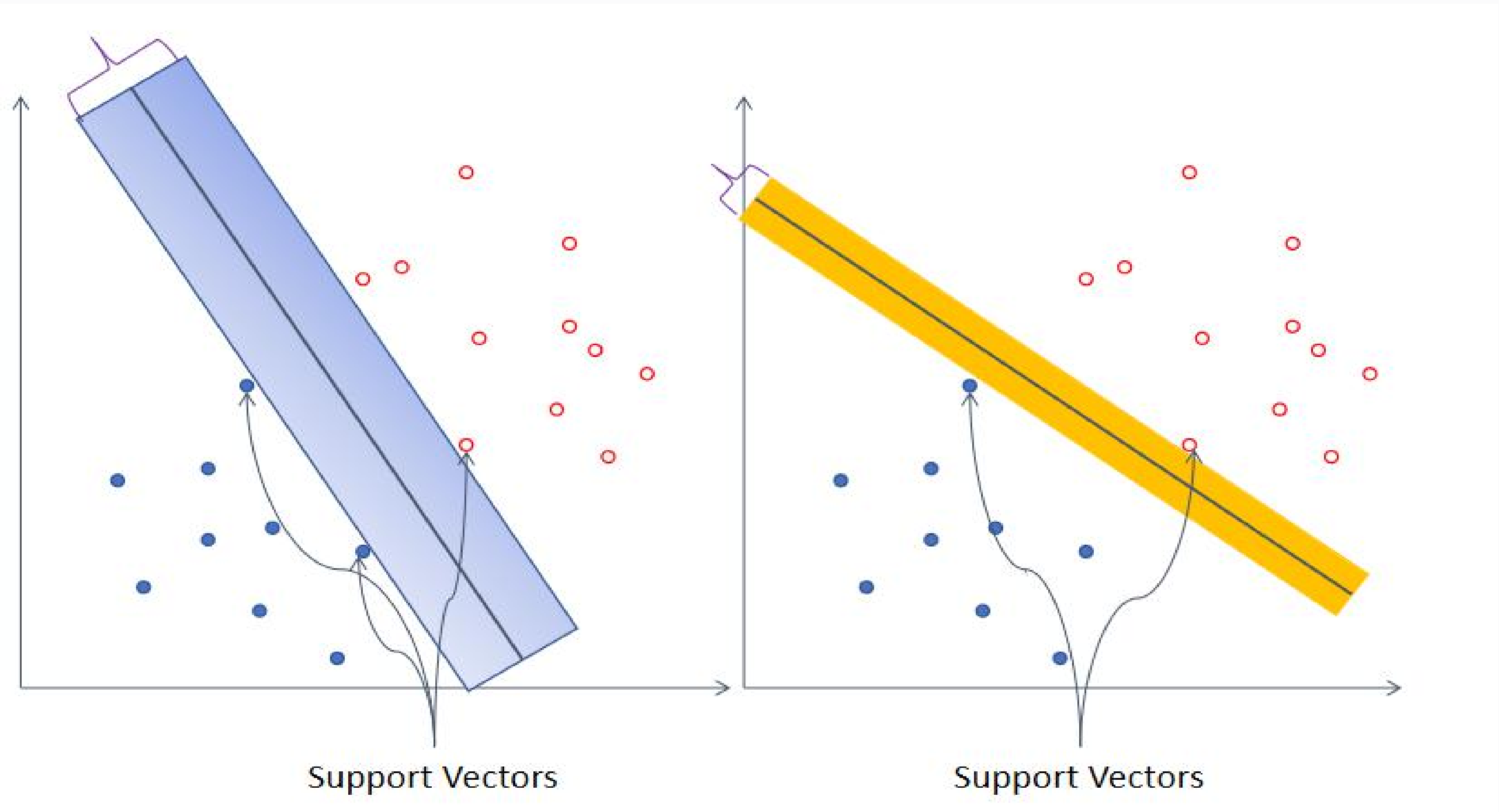
线性可分支持向量机与硬间隔最大化



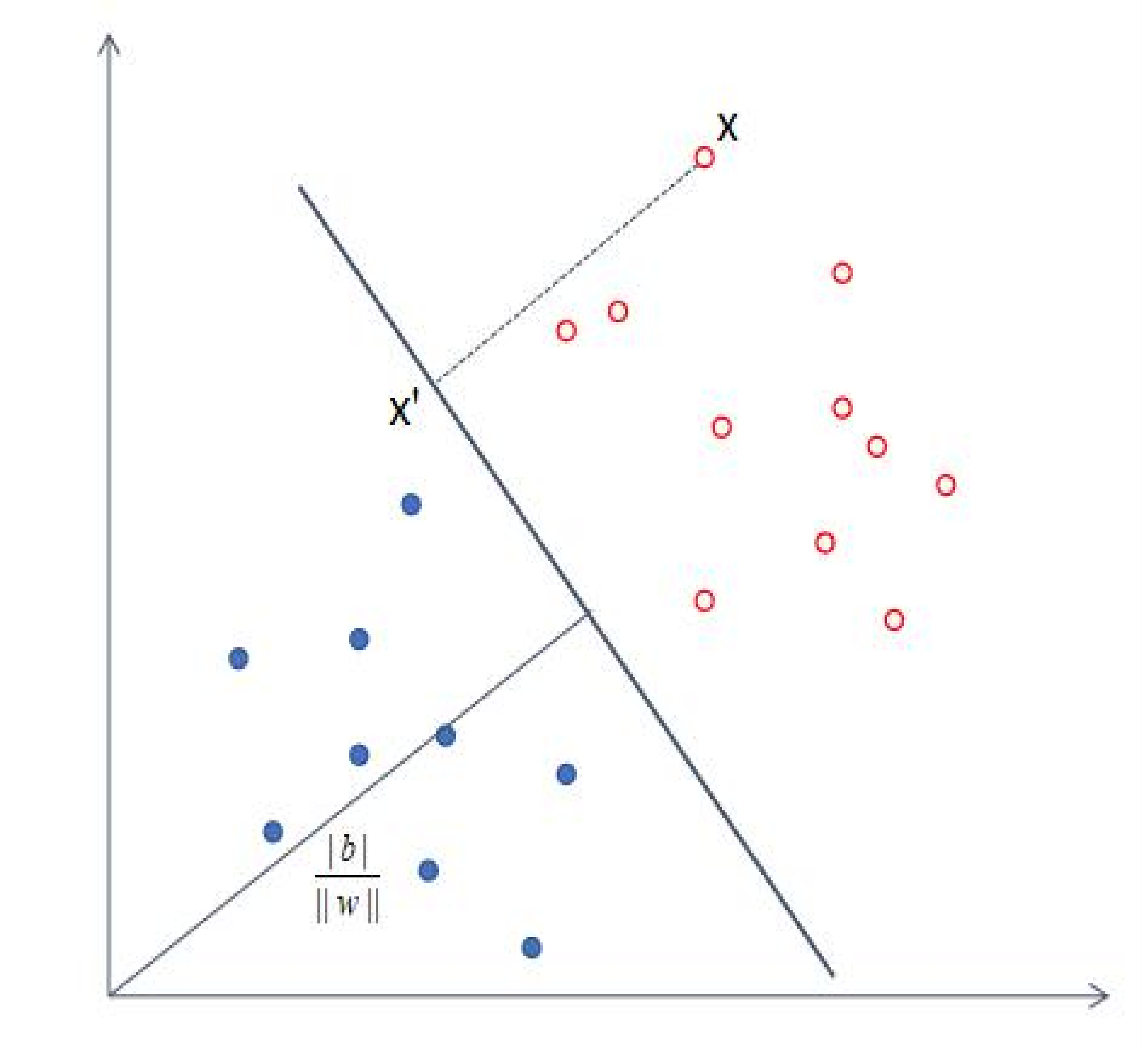
超平面选择



Margins



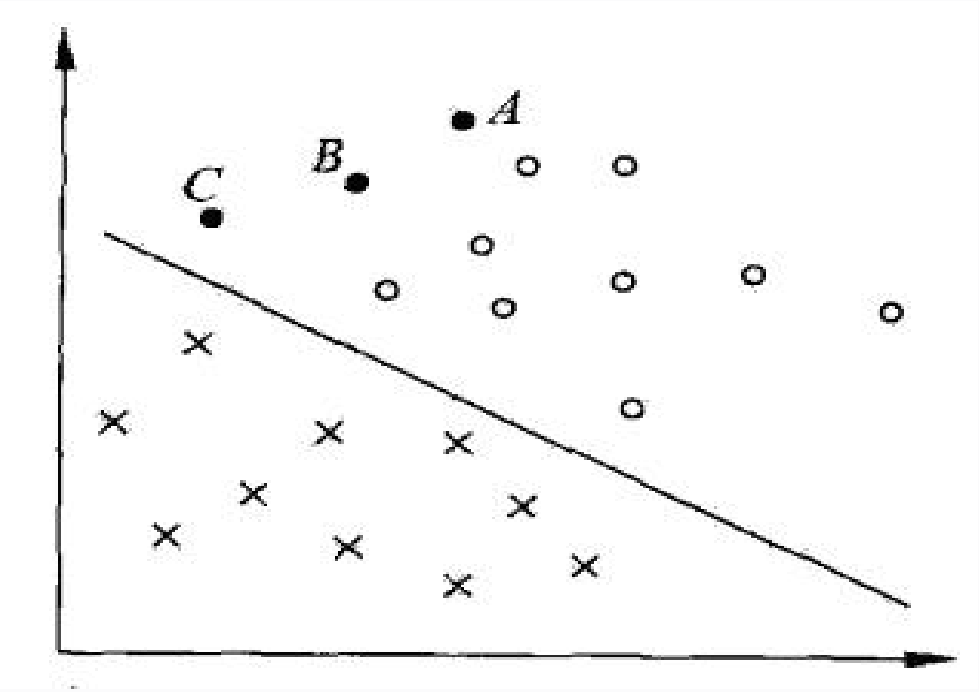
点到超平面的距离



$$g(x) = w \cdot x + b$$

函数间隔和几何间隔

- 点到分离超平面的远近 $|w \cdot x + b|$ 预测的确信程度
- $w \cdot x + b$ 的符号与类标记y的符号是否一致 表示分类是否正确
- 所以：$y(w \cdot x + b)$ 表示分类的正确性和确信度



函数间隔和几何间隔

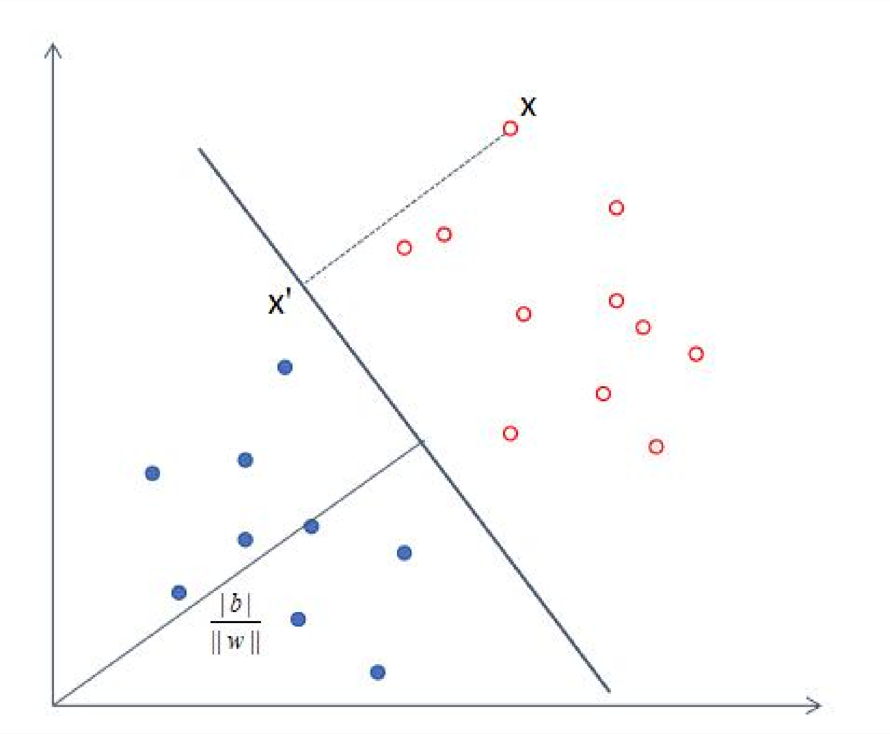
- 函数间隔
  - 样本点的函数间隔

$$\hat{\gamma}_i = y_i(w \cdot x_i + b$$

  - 训练数据集的函数间隔

$$\hat{\gamma} = \min_{i=1,\cdots,N} \hat{\gamma}_i$$

  - 表示分类预测的正确性和确信度
  - 当成比例改变w和b

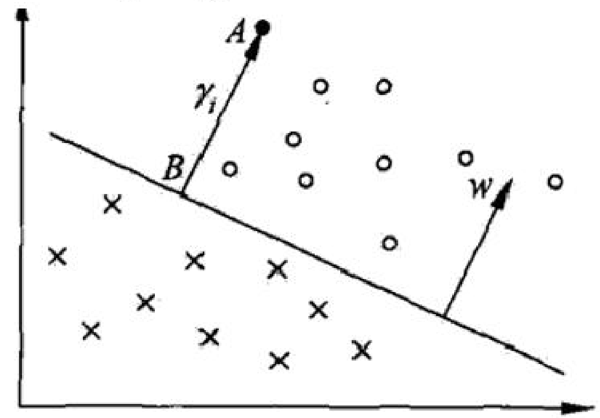


函数间隔和几何间隔

- 几何间隔
  - 样本点的几何间隔：正例和负例

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right) \qquad \gamma_i = -\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right)$$

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right)$$



函数间隔和几何间隔

- 几何间隔
  - 对于给定的训练数据集T和超平面(w, b)

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right)$$

  - 训练数据集的几何间隔 $\gamma = \min_{i=1,\cdots,N} \gamma_i$
  - 即

$$\gamma_i = \frac{\hat{\gamma}_i}{\|w\|}$$

$$\gamma = \frac{\hat{\gamma}}{\|w\|}$$

间隔最大化

- 最大间隔分类超平面

$$\max_{w,b} \quad \gamma$$

$$\text{s.t.} \quad y_i\left(\frac{w}{\|w\|}\cdot x_i+\frac{b}{\|w\|}\right)\geqslant\gamma, \quad i=1,2,\cdots,N$$

- 根据几何间隔和函数间隔的关系

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\|w\|}$$

$$\text{s.t.} \quad y_i(w\cdot x_i+b)\geqslant\hat{\gamma}, \quad i=1,2,\cdots,N$$

- 考虑
  - 可以取 $\hat{\gamma}=1$
  - 最大化 $\frac{1}{\|w\|}$ 和最小化 $\frac{1}{2}\|w\|^2$ 等价

间隔最大化

- 线性可分支持向量机学习的最优化问题

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

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i = 1,2,\cdots,N$$

- 凸二次规划(convex quadratic programming)

线性可分支持向量机学习算法

- 输入：线性可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$
  $x_i \in \mathcal{X}=\mathbf{R}^n \qquad y_i \in \mathcal{Y}=\{-1,+1\}$， $i=1,2,\cdots,N$
- 输出：最大间隔分离超平面和分类决策函数
- 1、构造并求解约束最优化问题
- $$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$ 求得w*和b*
  $$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i=1,2,\cdots,N$$
- 2、得到分离超平面 $w^* \cdot x + b^* = 0$
- 分类决策函数 $f(x)=\operatorname{sign}(w^* \cdot x + b^*)$

线性支持向量机与软间隔最大化

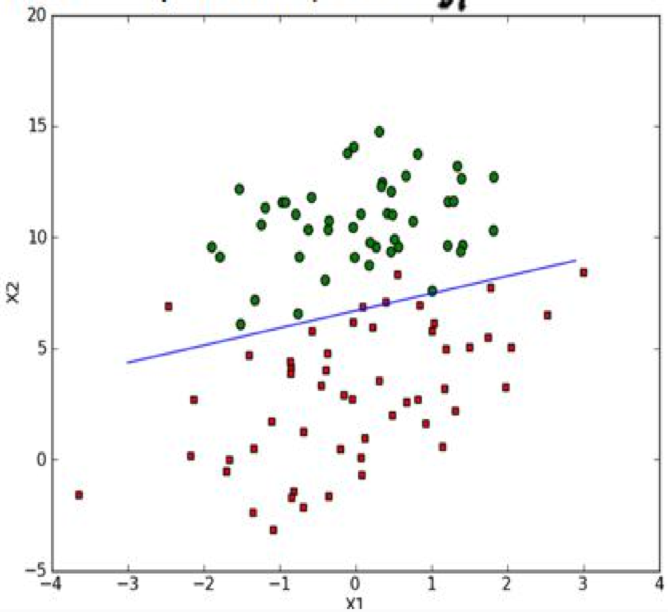
- 训练数据中有一些特异点（outlier），不能满足函数间隔大于等于1的约束条件。
- 解决方法：对每个样本点(x_i,y_i)引进一个松弛变量 $\xi_i \geq 0$
- 使得函数间隔加上松弛变量

大于等于1，约束条件变为：

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

目标函数变为： $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

C > 0 为惩罚参数



线性支持向量机与软间隔最大化

- 线性不可分的线性支持向量机的学习问题:

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

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant 1-\xi_i, \quad i=1,2,\cdots,N$$

$$\xi_i \geqslant 0, \quad i=1,2,\cdots,N$$

- 可证明w的解是唯一的，b不是，
- 设该问题的解是w*,b*,可得到分离超平面和决策函数

$$w^* \cdot x + b^* = 0$$

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

线性支持向量机与软间隔最大化

- 原始问题的拉格朗日函数:

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

- 其中: $\alpha_i \geqslant 0, \mu_i \geqslant 0$
- 对偶问题是拉格朗日函数的极大极小问题
- 首先求 $L(w,b,\xi,\alpha,\mu)$ 对w,b,ξ的极小，由

$$\nabla_w L(w,b,\xi,\alpha,\mu) = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0 \qquad w = \sum_{i=1}^{N}\alpha_i y_i x_i$$

- $$\nabla_b L(w,b,\xi,\alpha,\mu) = -\sum_{i=1}^{N}\alpha_i y_i = 0$$ 得: $$\sum_{i=1}^{N}\alpha_i y_i = 0$$ 带入

$$\nabla_{\xi_i} L(w,b,\xi,\alpha,\mu) = C - \alpha_i - \mu_i = 0 \qquad C - \alpha_i - \mu_i = 0$$

线性支持向量机与软间隔最大化

- 得：
$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

- 再对 $\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求α的极大，得到对偶问题：

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$C - \alpha_i - \mu_i = 0 \quad \Rightarrow \quad 0 \leqslant \alpha_i \leqslant C$$

$$\alpha_i \geqslant 0$$

$$\mu_i \geqslant 0, \quad i = 1,2,\cdots,N$$

线性支持向量机与软间隔最大化

- 原始问题 3 的对偶问题：

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

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$0 \leqslant \alpha_i \leqslant C, \quad i = 1,2,\cdots,N$$

4

- 定理：设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$ 是对偶问题 4 的一个解，若存在α*的一个分量 α_j^*，$0 < \alpha_j^* < C$，则原始问题的解w*,b*

$$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i \qquad b^* = y_j - \sum_{i=1}^{N} y_i \alpha_i^* (x_i \cdot x_j)$$

巴拿赫(Banach)空间

- 如果赋范线性空间(*X*, ||.||)是完备的，则称(*X*, ||.||)是Banach空间。

  - 例子
    - n 维Euclid空间$\boldsymbol{R}^n$是Banach空间
    - $L^p[a,b]$ $(p\geqslant 1)$是Banach空间，定义范数

$$\|x\| = \left[\int_a^b |x(t)|^p \,\mathrm{d}t\right]^{1/p}, p \geq 1$$

巴拿赫(Banach)空间

- 如果赋范线性空间(***X***, ||.||)是完备的，则称(***X***, ||.||)是Banach空间。
  - 例子：$C^k[a,b]$是Banach空间
    - $C^k[a,b]$表示定义在区间$[a,b]$上k阶连续可导的函数全体. 在$C^k[a,b]$定义范数

$$\|x\| = \sum_{j=0}^{k} \max\left|x^{(j)}(t)\right|, x^{(0)}(t) = x(t) \in C[a,b]$$

核函数在支持向量机的应用

- 注意到:
- 线性支持向量机对偶问题中，无论是目标函数还是决策函数都只涉及输入实例和实例之间的内积。
- 目标函数中的内积 $x_i \cdot x_j$ 用核函数 $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 代替，目标函数:

$$W(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i$$

- 决策函数: $f(x) = \mathrm{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i \phi(x_i) \cdot \phi(x) + b^*\right) = \mathrm{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i K(x_i, x) + b^*\right)$

常用核函数

- 1、多项式核函数（Polynomial kernel function)

$$K(x,z)=(x\cdot z+1)^p$$

- 对应的支持向量机为P次多项式分类器，分类决策函数:

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N_s}a_i^* y_i (x_i\cdot x+1)^p+b^*\right)$$

- 2、高斯核函数 （Gaussian Kernel Function)

$$K(x,z)=\exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$$

- 决策函数:

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N_s}a_i^* y_i \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)+b^*\right)$$

常用核函数

- 3、字符串核函数:
- 考虑一个有限字符表∑，字符串s是从∑中取出的有限字符的序列，包括空字符。s长度用|s|表示，元素记作 $s(1)s(2)\cdots s(|s|)$ 。两串字符s和t的连接记作st。所有长度为n的字符串的集合记作 Σ^n ，所有字符串记作 $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$
- 考虑s的子串u，给定一个指标序列 $i = (i_1, i_2, \cdots, i_{|u|})$，$1 \leqslant i_1 < i_2 < \cdots < i_{|u|} \leqslant |s|$；
- u定义为 $u = s(i) = s(i_1)s(i_2)\cdots s(i_{|u|})$ ，长度记作 $l(i) = i_{|u|} - i_1 + 1$
- 如果i是连续的，则 $l(i) = |u|$；，否则 $l(i) > |u|$。

非线性支持向量机学习算法

- 输入：线性不可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

  $x_i \in \mathcal{X}=\mathbf{R}^n \qquad y_i \in \mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$
- 输出：分类决策函数
- 1、选取适当的核函数和参数C，构造最优化问题：

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

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$0 \leqslant \alpha_i \leqslant C, \quad i=1,2,\cdots,N$$

5

求得最优解：$\alpha^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_N^*)^{\mathrm{T}}$

非线性支持向量机学习算法

- 2、并选择α*，适合条件 $0 < \alpha_j^* < C$ ，计算

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i \cdot x_j)$$

- 3、构造决策函数

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x \cdot x_i) + b^* \right)$$

- 当K(x,z)是正定核函数时，5 是凸二次规划问题，解是存在的。

谢谢各位大佬

感谢群里的大佬，谢谢闫老师

2020.2.25