支持向量机的原理

分享人：张召凯

问题

1 如何用支持向量机将一组正负样本分离

2 支持向量机和深度学习的关系

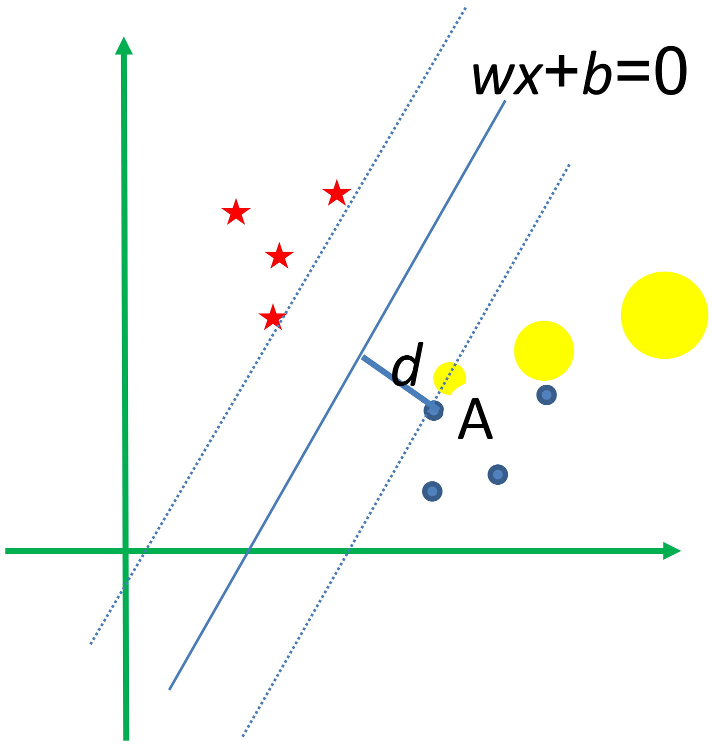
目　　录
CONTENTS

线性可分支持向量机定义：

训练数据线性可分时，通过硬间隔最大化，学习到的线性分类器。

考虑右图二维特征空间中的分类问题

- ★ 代表正样本
- ● 代表负样本



A表示一个距离超平面最近的样本，距离为d，最大化d。

LOGO

可以表示为下面约束最优化问题

$$\begin{aligned} \max_{w,b} \quad & d \\ s.t. \quad & d_i \geq d \quad , \quad i = 1, 2, \ldots, N \end{aligned}$$

解上述最优化问题，我们需要引入三个概念

LOGO

点到平面的距离公式

点x_0到超平面$wx+b=0$的距离公式

$$d=\frac{|wx_0+b|}{\|w\|}$$

几何间隔

几何间隔表示为γ，则样本点(x_i,y_i)的几何间隔为

$$\gamma_i=\frac{y_i(wx_i+b)}{\|w\|}$$

函数间隔

函数间隔表示为$\hat{\gamma}$,则样本点(x_i,y_i)的几何间隔为

$$\hat{\gamma}_i=y_i(wx_i+b)$$

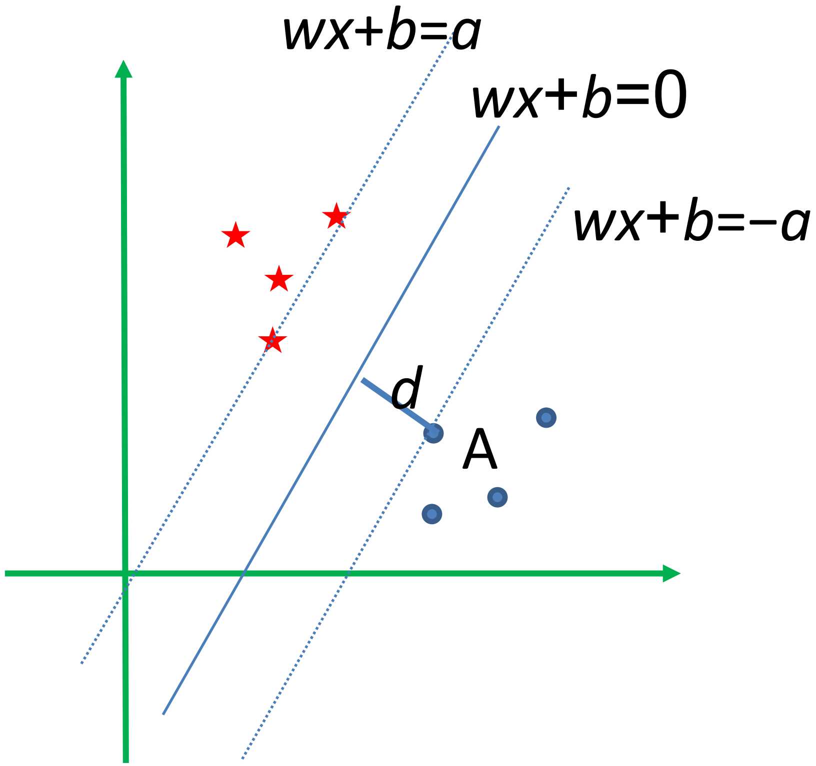
假定超平面为*wx*+*b*=0，那距离为*d到底是多少？*

我们假定间隔边界所在的平面为wx+b= ±*a*，*而A样本点位于边界之上，*

*我们称之为**支持向量**。*

d 可以表示为

$$d = \frac{a}{\|w\|}$$



LOGO

约束最优化问题可以转化成下述问题

$$\max_{w,b} \frac{a}{\|w\|}$$

$$s.t. \quad \frac{y_i(wx_i+b)}{\|w\|} \geq \frac{a}{\|w\|} \quad ,i=1,2,\ldots,N$$

为什么a可以用1代替

最常见的线性可分支持向量机的形式

$$\max_{w,b} \frac{1}{\|w\|}$$

$$s.t. \quad y_i(wx_i+b) \geq 1 \quad ,i=1,2,\ldots,N$$

对于任意的函数间隔±*a*,在x不变的情况的下，我们都可以找到一个因子λ，使得函数间隔变为1，即$\lambda wx+\lambda b=1$

对于我们最终想要的超平面$wx+b=0$来说， *如果得到*$\lambda wx+\lambda b=0$的话，是一样的。

在变换前后，支持向量不变，支持向量到超平面的距离也不变。

LOGO

目标函数的替换

$$\max_{w,b} \quad \frac{1}{\|w\|}$$
$$s.t. \quad y_i(wx_i+b) \geq 1 \quad ,i=1,2,\ldots,N$$

↓

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$s.t. \quad y_i(wx_i+b) \geq 1 \quad ,i=1,2,\ldots,N$$

这个替换和我们本科是用拉格朗日条件极值解题做的替换一样——方便计算

LOGO

构造拉格朗日函数

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

$$\begin{aligned} s.t. \quad & \alpha_i \geq 0 \\ & y_i(w\cdot x_i+b)-1\geq 0 \qquad i=1,2,\ldots,N \end{aligned}$$

上式中第二项恒大于等于零，所以上述函数的最大值存在的情况下，就为第一项的值

优化问题转化为

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

L是凸函数，属于凸函数优化问题。

此处的凸函数定义与同济高数书上的定义是反的

LOGO

拉格朗日对偶问题

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

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

当w,α*满足KKT条件时，原始问题的解和对偶问题的解相同*

$$\begin{aligned}
&\nabla_w L(w^*,b^*,\alpha^*)=0\\
&\nabla_b L(w^*,b^*,\alpha^*)=0\\
&\alpha^*(y_i(w^*\cdot x_i+b^*)-1)=0\\
&y_i(w^*\cdot x_i+b^*)-1\geq 0\\
&\alpha^*\geq 0 \qquad i=1,2,\ldots,N
\end{aligned}$$

LOGO

对变量w, b, α求偏导，并令偏导为0可得：

$$\omega = \sum_{i=1}^{m} \alpha_i y_i x_i$$

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

将结果代入

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

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

$$\alpha_i \geq 0 \qquad i = 1, 2, \ldots, N$$

根据KKT条件，可以得到原始最优化问题的解

$$w^* = \sum_{i=1}^{N} \alpha_i y_i x_i$$

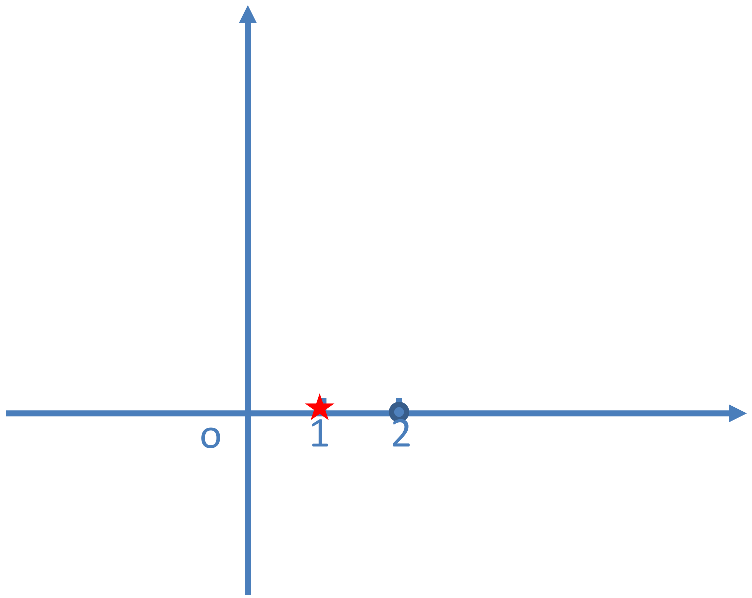
$$b^* = y_j - \sum_{i=1}^{N} \alpha_i y_i (x_i \cdot x_j)$$

（xj，yj）是α_j不为0的样本

分类决策函数

$$f(x) = sgn(\sum_{j=1}^{N} \alpha_i y_i (x_i \cdot x) + b^*)$$

如下图所示训练集，正样本点（1，0），负样本点（2，0），试求最大间隔分离超平面。

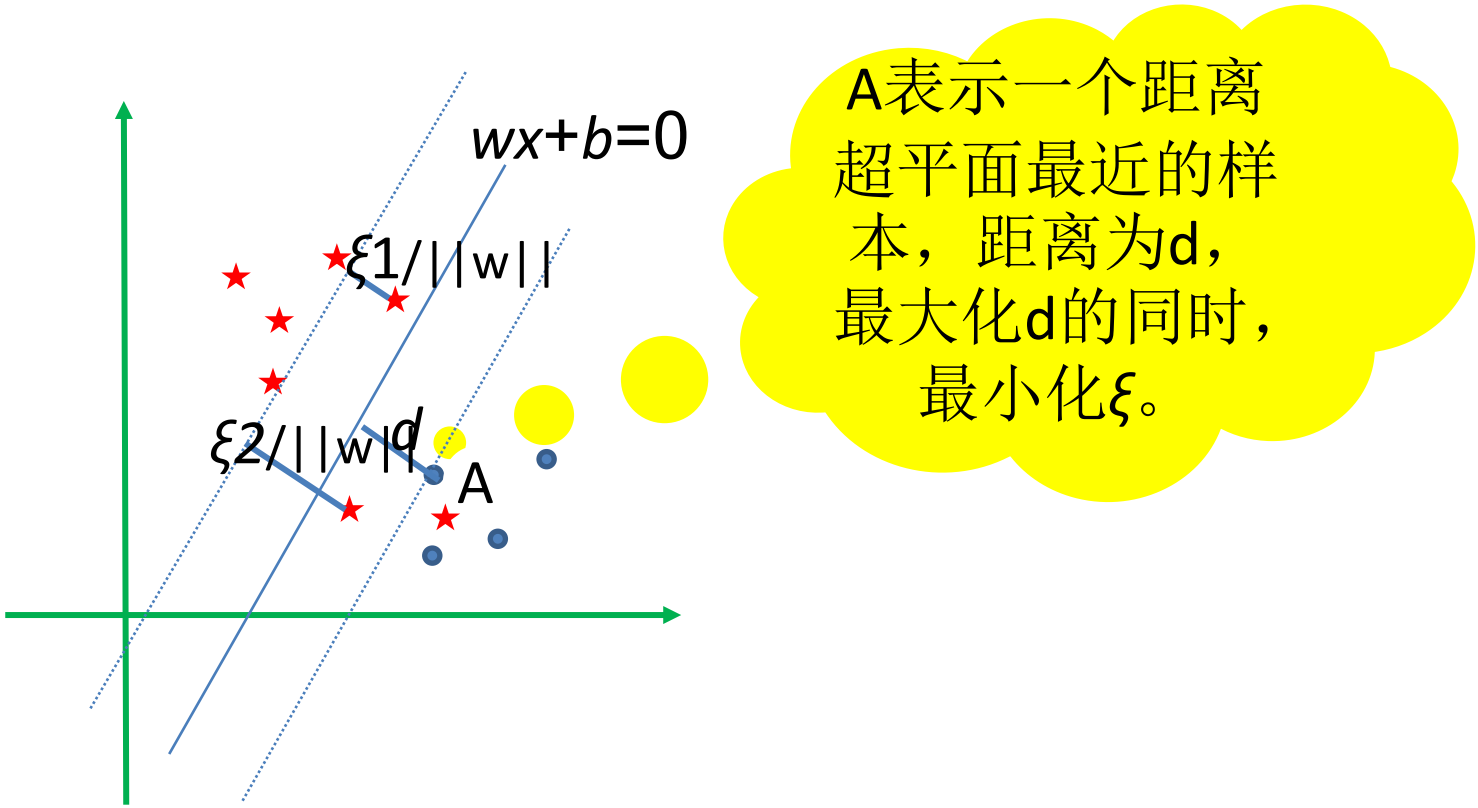


线性支持向量机定义：

在训练数据线性不可分时，通过软间隔最大化，学习到的线性分类器。

考虑右图二维特征空间中的分类问题

- 代表正样本
- 代表负样本



LOGO

约束最优化问题

$$\min \frac{\|w\|^2}{2} + C\xi_i$$

$$s.t. \quad y_i(wx_i + b) \geq 1 - \xi_i \quad i = 1, 2, \ldots, N$$

$$\xi_i \geq 0 \quad i = 1, 2, \ldots, N$$

松弛变量ξ：

惩罚参数C：

最小化目标函数的两层含义：
1、间隔尽量大
2、误分类点的个数尽量小

LOGO

构造拉格朗日函数

$$L(w,b,\alpha,\mu)=\frac{\|w\|^2}{2}-\sum_{i=1}^{N}\alpha_i\{y_i(w\cdot x_i+b)-1+\xi_i\}-\sum_{i=1}^{N}\mu_i\xi_i$$
$$s.t.\quad \alpha_i\geq 0\quad \mu_i\geq 0\quad i=1,2,\ldots,N$$

转化为拉格朗日对偶问题，分别对*w, b, α, μ求偏导，并令偏导数为0，得*

$$\max_{\alpha}-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i$$
$$s.t.\quad \sum_{i=1}^{N}\alpha_i y_i=0$$
$$\boxed{C-\alpha_i-\mu_i=0}$$
$$\alpha_i\geq 0$$
$$\mu_i\geq 0\quad i=1,2,\ldots,N$$

LOGO

消去μ

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

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

$$\boxed{0 \le \alpha_i \le C} \quad i = 1, 2, \ldots, N$$

形式上和线性可分支持向量机差不多，只是求b时，需要选取0<α<C*对应的样本进行计算*

支持向量	
线性可分支持向量机	线性支持向量机
● 间隔边界上的向量	● 间隔边界上的向量（$0<\alpha<C$，此时$\xi=0$，μ为任意值） ● 间隔边界与超平面之间的向量（$\alpha=C$，此时$\mu=0$，$0<\xi<1$） ● 超平面上的向量（$\alpha=C$，此时$\mu=0$，$\xi=1$） ● 分离超平面误分类的向量（$\alpha=C$，此时$\mu=0$，$\xi>1$）

LOGO

线性支持向量机的另一种形式

$$\min_{w,b} \quad \sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ + \lambda\|w\|^2$$

$$\begin{aligned} &\min \frac{\|w\|^2}{2} + C\xi_i \\ s.t. \quad & y_i(wx_i+b) \geq 1-\xi_i \qquad i=1,2,\ldots,N \\ & \xi_i \geq 0 \qquad i=1,2,\ldots,N \end{aligned}$$

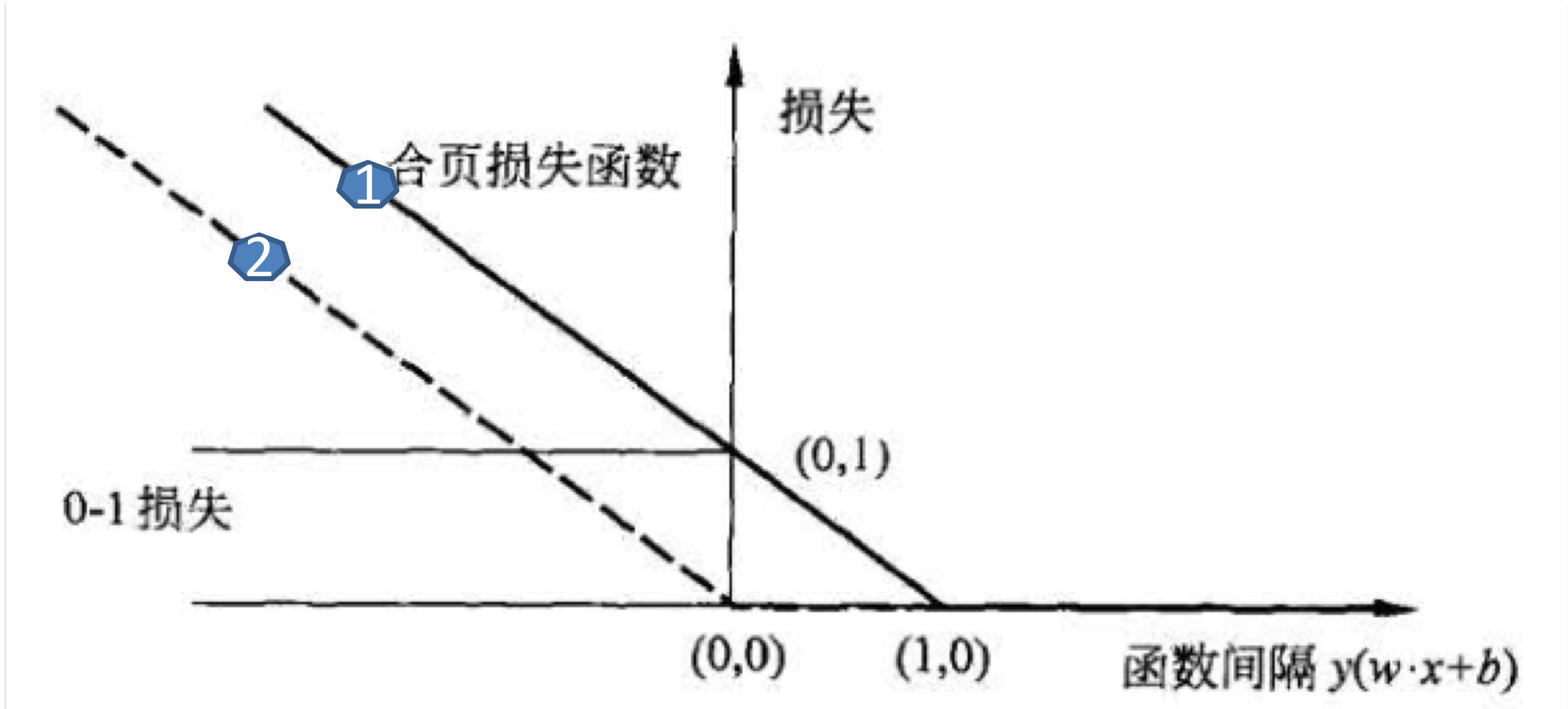
合页损失函数(hinge loss funtion)

$$L(y(wx+b)) = [1-y(wx+b)]_+$$

取正值函数

$$[z]_+ = \begin{cases} z, z>0 \\ 0, z\leq 0 \end{cases}$$

合页损失函数的应用



① 为线性支持向量机的损失函数
② 为感知机的损失函数

	线性可分支持向量机	线性支持向量机	深度学习
是否有损失	无	有	有
损失函数	无	合页损失函数+L2正则化	最小均方差
学习方法	凸二次规划	凸二次规划	梯度下降
适用范围	线性可分	近似线性可分	线性和非线性都适用
数据集数量依赖	低	低	高

LOGO

参考

1、统计学习方法-李航
2、 https://www.cnblogs.com/xxrxxr/p/7536131.html
3、https://www.jianshu.com/p/fe14cd066077

感谢

贾鑫康推荐编写数学公式的工具

Vscode+ Markdown All in One（plugin）

谢谢