# Chapter7支持向量机1Support Vector Machine（SVM）

1. 线性支持向量机
2. 核支持向量机
3. 序列最小优化算法

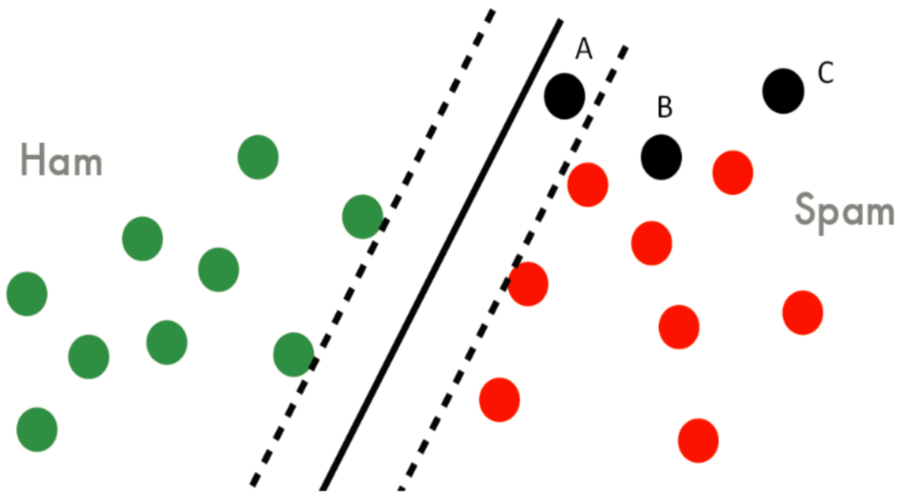
## 超平面选择

1. 超平面选择：best hyperplane是最大化两个类别边距的那种方式；超平面对每个类别最近的元素最远

   > 💡 **SVM和logistic回归的区别**
   >
   > SVM更关心的是靠近中间分割线的点，让他们尽可能地远离中间线，而不是在所有点上达到最优，因为那样的话，要使得一部分点靠近中间线来换取另外一部分点更加远离中间线。
   >
   > 因此支持向量机和和逻辑斯蒂回归的不同点，一个是考虑局部（不关心已经确定远离的点，更考虑靠近中间分割线的点），一个是考虑全局（已经远离的点可能通过调整中间线使其能够更加远离）

2. 间隔Margin

   点到分离超平面的距离反映了预测的置信度




# 二分类问题

## 线性分类器

■**样本** $(\boldsymbol{x}, y), y \in \{-1,1\}$

■线性分类器:

$$f_{w,b}(\boldsymbol{x}) = g(\boldsymbol{w}^T \boldsymbol{x} + b)$$

$$g(z) = \begin{cases} 1, & \text{如果 } z \geq 0 \\ -1, & \text{如果} z < 0 \end{cases}$$

■超平面: $\boldsymbol{w}^T \boldsymbol{x} + b = 0$ .

### 函数间隔(Functional Margin)

■对于一个训练样本 $(\boldsymbol{x}^i, y^i)$ ，它到 $(\boldsymbol{w}, b)$ 确定的超平面的函数间隔为：

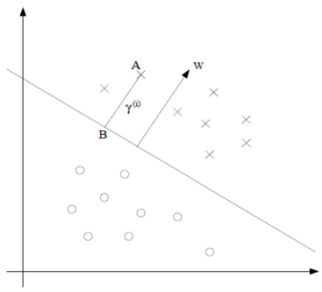
$$\hat{\gamma}^i = y^i(\boldsymbol{w}^T\boldsymbol{x}^i + b)$$

- $y^i = 1$ ， $\boldsymbol{w}^T\boldsymbol{x}^i + b$ 是一个大的正数.
- $y^i = -1$ ， $\boldsymbol{w}^T\boldsymbol{x}^i + b$ 是一个比较小的负数.
- $y^i(\boldsymbol{w}^T\boldsymbol{x}^i + b) > 0$ ，模型对样本的预测是正确的.
- 大的函数间隔 → 确信正确的预测

■对于训练数据集 $S = \{(\boldsymbol{x}^i, y^i), i = 1, \ldots, N\}$ ，它的函数间隔定义为所有样本中最小的那个

$$\hat{\gamma} = \min_i \hat{\gamma}^i, i = 1, 2, \ldots, N$$

**几何间隔(Geometric Margin)**




■点B： $\boldsymbol{x}^i - \gamma^i \boldsymbol{w} / \|\boldsymbol{w}\|_2$

■点B在决策边界上：

$$\boldsymbol{w}^T(\boldsymbol{x}^i - \gamma^i \boldsymbol{w} / \|\boldsymbol{w}\|_2) + b = 0$$

■求解方程可得 $\gamma^i$ ：

$$\gamma^i = \frac{\boldsymbol{w}^T\boldsymbol{x}^i + b}{\|\boldsymbol{w}\|_2} = \left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}$$

■样本 $(\boldsymbol{x}^i, y^i)$ 的几何间隔为：

$$\gamma^i = y^i\left(\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}\right)$$

■训练数据集 $S=\{(\boldsymbol{x}^i, y^i), i=1,\ldots,N\}$ 关于判别界面（参数为$(\boldsymbol{w},b)$））的几何间隔：

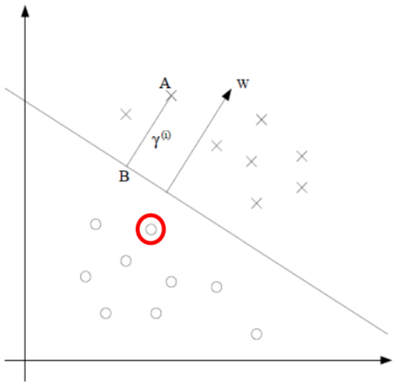
$$\gamma = \min_i \gamma^i, i=1,2,\ldots,N$$

■函数间隔与几何间隔的关系：

$$\gamma^i = \frac{\hat{\gamma}^i}{\|\boldsymbol{w}\|_2}, \quad \gamma = \frac{\hat{\gamma}}{\|\boldsymbol{w}\|_2}.$$

•如果 $\|\boldsymbol{w}\|_2 = 1$ ， 二者相等。

■**几何间隔具有不变性.**




> 💡 因为函数间隔是我们定义的，在定义的时候就有几何间隔的色彩。几何间隔最大的优势就是不管将w和b扩大几倍，几何间隔都没有影响。

## 最优间隔分类器

给定一个训练集，一个自然的想法是试图找到一个使几何间隔最大化的决策边界，这表示对训练集的有可信的预测并且对训练数据的良好“拟合”。

**间隔最大化**

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

$$s.t. \ \ y^i\left(\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}\right)^T \boldsymbol{x}^i + \frac{b}{\|\boldsymbol{w}\|_2}\right) \geq \gamma, i=1,\ldots,N$$

将问题转换为

$$\max_{\gamma, \boldsymbol{w}, b} \frac{\hat{\gamma}}{\|\boldsymbol{w}\|_2}$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T \boldsymbol{x}^i + b\right) \geq \hat{\gamma}, i = 1, \ldots, N$$

给定 $\hat{\gamma} = 1$简化问题

$$\min_{\boldsymbol{w}, b} \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T \boldsymbol{x}^i + b\right) \geq 1, i = 1, \ldots, N$$

**求解：使用二次规划(QP)方法求解**

**二次规划**(QP, Quadratic Programming)定义:目标函数为二次函数，约束条件为线性约束，属于最简单的一种非线性规划。

# 线性SVM

## 基本定义

1. 基本定义

■输入：线性可分的训练数据集 $S = \{(\boldsymbol{x}^i, y^i), i = 1, \ldots, N\}$

■输出：判别函数及决策/判别界面

通过求解如下最优化问题来得到最优分类器的参数 ($w$*,$b$)

•

$$\min_{w,b} \frac{1}{2}\|w\|_2^2$$
$$s.t.\ \ y^i\left(w^T x^i + b\right) \geq 1, i = 1, \ldots, N$$

•分离超平面: $(w^*)^T x + b^* = 0$

•判别函数： $f_{w,b}(x) = sign((w^*)^T x + b^*)$

**理论支持：对于线性可分的训练数据集，最大间隔分类器存在且唯一。**

2. 支持向量和间隔

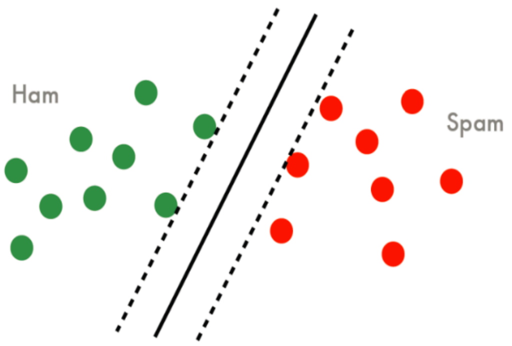
■**支持向量：**距分离超平面最近的训练样本。

$$y\left(w^T x + b\right) = 1$$

■函数间隔: $\hat{\gamma} = 1$

■几何间隔: $\gamma = \frac{1}{\|w\|_2}$

■间隔: $\frac{2}{\|w\|_2}$




## 拉格朗日对偶→求解

原始的最优化问题：

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0, i = 1, 2, \ldots, m \\ & h_j(x) = 0, j = 1, 2, \ldots, p \end{aligned}$$

💡 原始问题通常会带有很多约束条件，这样直接求解原始问题往往是很困难的，于是考虑将原始问题转化为它的对偶问题，通过求解它的对偶问题来得到原始问题的解。**对偶性**（**Duality**）是凸优化问题的核心内容。

将原始问题的拉格朗日函数定义为：

$$L(x,\lambda,v)=f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{j=1}^{p}v_j h_j(x)$$

其中 $x\in R^n,\lambda\in R^m,v\in R^p$ 。

💡 拉格朗日函数L可以看成关于x的函数，它实际上是**对原始问题中目标函数与约束条件**进行**线性加权**。目标函数的权系数是1，约束条件的权系数是 $\lambda_i$或 $v_i$。若将L看成关于 $\lambda_i$， $v_i$的函数，则其余部分可以看成常数，L为一个关于 $\lambda_i$和 $v_i$仿射函数。

💡 https://blog.csdn.net/frostime/article/details/90291392

- 简单情形

## ■等式约束的最优化问题:

$$\min_{w} f(w)$$
$$s.t.\quad h_i(w)=0,\ i=1,\dots,l$$

•拉格朗日函数：

$$L(w,\beta)=f(w)+\sum_{i=1}^{l}\boxed{\beta_i}h_i(w)$$

拉格朗日乘子

## ■求解：偏导数为0

$$\frac{\partial L}{\partial w_i}=0,\quad \frac{\partial L}{\partial \beta_i}=0$$

- 一般情形

**基本定义**

### ■约束条件中含有等式及不等式约束时:

$$\boxed{\begin{aligned}&\min_{w} f(w)\\ &s.t.\ \ g_i(w)\le 0,\ i=1,\dots,k\\ &\qquad h_i(w)=0,\ i=1,\dots,l\end{aligned}}$$

原问题
Primary Problem

### ■广义拉格朗日函数:

非负

$$L(w,\alpha,\beta)=f(w)+\sum_{i=1}^{k}\boxed{\alpha_i}g_i(w)+\sum_{i=1}^{l}\beta_i h_i(w)$$

拉格朗日乘子

**原始问题定义：Primitive Problem→ $\theta_p$**

### ■定义如下最优化问题:

$$\theta_P(w)=\max_{\alpha,\beta:\alpha_i\ge 0} L(w,\alpha,\beta)$$

■如果满足所有约束条件:

$$\theta_P(\boldsymbol{w}) = f(w)$$

■分析下面的最优化问题:

$$\min_{w} f(w) = \min_{w} \theta_P(\boldsymbol{w}) = \min_{w} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

■原始问题（Primitive problem）：

$$\boxed{\begin{aligned} &\min_{\boldsymbol{w}} f(\boldsymbol{w}) \\ &s.t.\ \ g_i(\boldsymbol{w}) \leq 0,\ i = 1,\dots,k \\ &\qquad h_i(\boldsymbol{w}) = 0,\ i = 1,\dots,l \end{aligned}}$$

■定义该问题的最优值：

$$p^* = \min_{\boldsymbol{w}} \theta_P(\boldsymbol{w})$$ 原问题的最优值

$$p^* = min_w \theta_p(w)$$

**对偶问题定义：Dual Problem**

■对偶问题（Dual problem）：

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_{\alpha,\beta:\alpha_i \geq 0} \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$\theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})$$

■对偶问题的最优值:

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta})$$

■原问题与对偶问题：

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} \min_{\boldsymbol{w}} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \min_{\boldsymbol{w}} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i \geq 0} L(\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta}) = p^*$$

**Dual Problem & Primitive Problem**

$$d^* = max_{\alpha,\beta;\alpha_i \geq 0} min_w L(w,\alpha,\beta) \leq min_w max_{\alpha,\beta;\alpha_i \geq 0} L(w,\alpha,\beta) = p^*$$

**那么什么情况下可以找到 $d^* = p^*$，即满足强对偶→ KKT 条件**

- 如果原问题是**凸优化问题**，且存在**严格满足约束条件**的$w$。

$$\min_w f(w)$$
$$s.t.\ g_i(w) \leq 0,\ i=1,\ldots,k$$
$$h_i(w) = 0,\ i=1,\ldots,l$$

$f(.)$、$g_i(.)$是凸函数，$h_i(.)$是仿射函数 affine function

- 则存在 $w^*, \alpha^*, \beta^*$，使得 $w^*$ 是原问题的解，$\alpha^*, \beta^*$ 是对偶问题的解，且满足：

$$p^* = d^* = L(w^*, \alpha^*, \beta^*)$$

- $w^*, \alpha^*, \beta^*$ 满足Karush-Kuhn-Tucker (KKT) 条件:

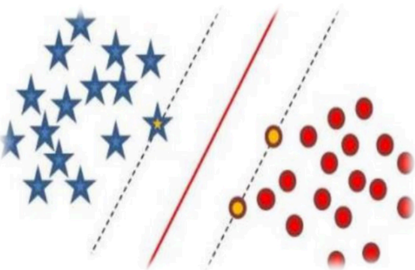
$$\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0,\ i=1,2,\ldots,M$$
$$g_i(w^*) \leq 0,\ i=1,\ldots,k$$
$$h_i(w^*) = 0,\ i=1,\ldots,l$$
$$\boxed{\alpha_i^* g_i(w^*) = 0,\ i=1,\ldots,k}$$
$$\alpha_i^* \geq 0,\ i=1,\ldots,k$$

KKT 对偶互补性

**最优间隔分类器：对偶解**

- 最优间隔分类器——原问题:

$$\min_{w,b} \frac{1}{2}\|w\|_2^2$$
$$s.t.\ y^i\left(w^T x^i + b\right) \geq 1, i=1,\ldots,N$$



- 约束条件:

$$g_i(w) = -y^i\left(w^T x^i + b\right) + 1 \leq 0,\ i=1,\ldots,N$$

- KKT对偶互补条件: $\alpha_i^* g_i(w^*) = 0,\ i=1,\ldots,k$

$$g_i(w^*) < 0 \Rightarrow \alpha_i^* = 0$$
$$\boxed{\alpha_i^* \neq 0 \Rightarrow g_i(w^*) = 0}$$

**支持向量**

- 支持向量的数量远小于训练样本的数目!

💡 **支持向量**（Support Vectors）是那些对决策边界起关键作用的训练样本。这些样本在训练过程中定义了分类超平面的位置，而**其他样本**则不会直接影响决策边界的选择。

因此，支持向量数量远小于训练样本数目。

### 对偶解

■拉格朗日函数:

$$L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|_2^2-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1\right]$$

■对偶解:

- 固定 $\boldsymbol{\alpha}$， 关于参数 $\boldsymbol{w}$ 和 $b$ 最小化 $L(\boldsymbol{w},b,\boldsymbol{\alpha})$ 得到 $\theta_D(\boldsymbol{\alpha})$

$$\theta_D(\boldsymbol{\alpha})=\min_{\boldsymbol{w},b}L(\boldsymbol{w},b,\boldsymbol{\alpha})$$

- 最大化 $\theta_D(\boldsymbol{\alpha})$ 得到对偶问题的最优值 $d^*$

■ 固定 $\boldsymbol{\alpha}$ ，关于参数 $\boldsymbol{w}$ 和 $b$ 最小化 $L(\boldsymbol{w},b,\boldsymbol{\alpha})$ 得到 $\theta_D(\boldsymbol{\alpha})$

- **求解 $\boldsymbol{w}$ 和 $b$** $\boxed{L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|_2^2-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1\right]}$

$$\nabla_{\boldsymbol{w}}L(\boldsymbol{w},b,\boldsymbol{\alpha})=\boldsymbol{w}-\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i=0,$$

$$\Downarrow$$

$$\boldsymbol{w}=\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i$$

$$\frac{\partial}{\partial b}L(\boldsymbol{w},b,\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i y^i=0,$$

- 代入拉格朗日函数

$$\theta_D(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$$

## ■对偶问题

$$\boxed{\begin{aligned}&\max_{\boldsymbol{\alpha}}\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j\\&s.t.\ \alpha_i\geq 0, i=1,..,N,\\&\sum_{i=1}^{N}\alpha_i y^i=0.\end{aligned}}$$

对偶解推导

下一步求解

■ 输入：线性可分的训练数据集 $S=\{(\boldsymbol{x}^i,y^i),i=1,\ldots,N\}$

■ 输出：分离超平面和判别函数.

- 通过求解对偶问题来得到最优解 $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_l^*)$

$$\max_{\boldsymbol{\alpha}}\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j$$

$$s.t.\ \alpha_i\geq 0,i=1,..,N,$$

$$\sum_{i=1}^{N}\alpha_i y^i=0.$$

- 得到原问题的最优解 $(\boldsymbol{w}^*,b^*)$

$$\boldsymbol{w}^*=\sum_{i=1}^{N}\alpha_i^*y^i\boldsymbol{x}^i,\ b^*=y^j-\sum_{i=1}^{N}\alpha_i^*y^i\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.\quad \alpha_j^*>0$$

•分离超平面: $(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*=0$ ，判别函数: $f_{\boldsymbol{w},b}(\boldsymbol{x})=(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*.$

其中，需要注意的：使用支持向量进行计算

$$\boldsymbol{w}^*=\sum_{i=1}^{N}\alpha_i^*y^i\boldsymbol{x}^i,\ b^*=y^j-\sum_{i=1}^{N}\alpha_i^*y^i\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.\qquad \alpha_j>0$$

■计算$\boldsymbol{w}^*$ 时 $b^*$ ，只需要利用 $\alpha_i>0$ 的那些样本(支持向量)来计算。

$$f_{\boldsymbol{w},b}(\boldsymbol{x})=(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*=\sum_{i=1}^{N}\alpha_i^*y^i\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}+b^*$$

■预测时是否需要所有的训练样本来计算?

■对偶形式：只需计算训练样本与输入特征的内积.

$$\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}=<\boldsymbol{x}^i,\boldsymbol{x}>$$

■**核技巧!**

## 软间隔Soft Margin分类器

解决训练数据线性不可分的情况

■允许一些样本（离群点或噪声样本）违反原来的不等式约束条件: $y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\geq 1$，但是数目要少。

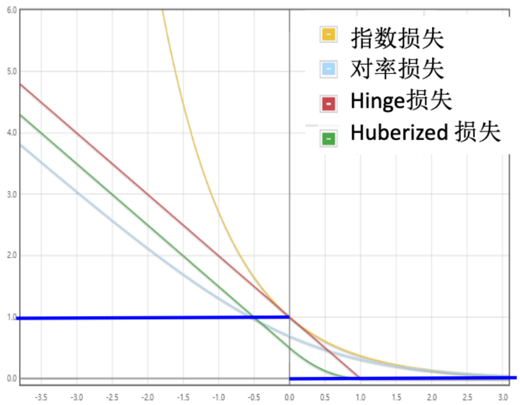
■ 目标函数:

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}I_{y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)<1}$$

■ 不易求解!

■ 替代损失:

0/1损失




hinge损失: $l_{hinge}$(z)=max(0,1-z)

指数损失: $l_{exp}$(z)= exp(-z)

Huberized损失: $l_{hub}$(z)=0.5-z, z<0; 0.5(1-z)$^2$,0 ≤ z ≤ 1 ; 0, z>1

对率损失(logistic): $l_{log}$(z)=log(1+exp(-z))

1. Hinge损失

■Hinge 损失:

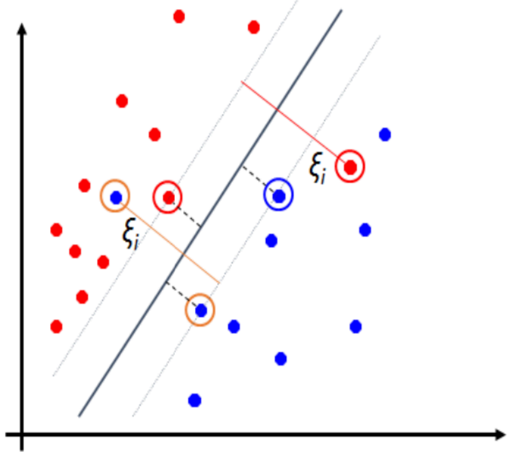
$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}\max\left\{0,1-y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\right\}$$

■松弛变量 $\xi_i$ :

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}\xi_i$$

$$s.t.\ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\geq 1-\xi_i,\ i=1,\ldots,N,$$

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




■软间隔分类器.

■凸二次规划问题。

为了描述这些错误，我们引入了**松弛变量** $\xi_i$来表示**第i个样本违反间隔的程度**，并引入一个**损失函数**来量化**这些错误**。

目标是**最小化间隔最大化和分类错误之间的折中。**

■拉格朗日函数 $L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\eta})=\frac{1}{2}\|\boldsymbol{w}\|_2^2+C\sum_{i=1}^{N}\xi_i$

$-\sum_{i=1}^{N}\alpha_i\left[y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)-1+\xi_i\right]-\sum_{i=1}^{N}\eta_i\xi_i.$

■固定 $\boldsymbol{\alpha}$、$\boldsymbol{\eta}$，关于参数$\boldsymbol{w}$ 、$b$ 和 $\xi$，最小化 $L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\eta})$，得到 $\theta_D(\boldsymbol{\alpha},\boldsymbol{\eta})$

$$\theta_D(\boldsymbol{\alpha},\eta)=\min_{w,b,\xi}L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\eta)\quad\left\{\begin{array}{l}\nabla_w L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\eta})=\boldsymbol{w}-\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i=0, \Rightarrow \boxed{\boldsymbol{w}=\sum_{i=1}^{N}\alpha_i y^i\boldsymbol{x}^i}\\ \frac{\partial}{\partial b}L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\eta})=\sum_{i=1}^{N}\alpha_i y^i=0,\\ \frac{\partial}{\partial \xi_i}L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\eta})=C-\alpha_i-\eta_i=0.\end{array}\right.$$

$$\theta_D(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$$

■最大化 $\theta_D(\boldsymbol{\alpha},\boldsymbol{\eta})$，得到最优值 $d^*$。

**■对偶问题**

$$\max_{\boldsymbol{\alpha}}\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}y^iy^j\alpha_i\alpha_j\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j$$

$$s.t.\ 0\le\alpha_i\le C, i=1,..,N,$$

$$\sum_{i=1}^{N}\alpha_i y^i=0.$$

■ 输入： 训练数据集 $S=\{(\boldsymbol{x}^i,y^i),i=1,\ldots,N\}$

■ 输出： 分类超平面与分类决策函数.

- 选择参数$C$， 并求解对偶问题，得最优解 $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_l^*)$

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y^i y^j \alpha_i \alpha_j \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j$$

$$s.t.\ 0 \le \alpha_i \le C, i=1,..,N,$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0.$$

- 求解最优的 $(\boldsymbol{w}^*,b^*)$ ：$\boldsymbol{w}^*=\sum_{i=1}^{N}\alpha_i^* y^i \boldsymbol{x}^i,\ b^*=y^j-\sum_{i=1}^{N}\alpha_i^* y^i\left(\boldsymbol{x}^i\right)^T\boldsymbol{x}^j.$ $\boxed{0<\alpha_j^*<C}$

•分类超平面: $(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*=0$ ，判别函数: $f_{\boldsymbol{w},b}(\boldsymbol{x})=(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*.$

> 💡 Different between the hinge and the Lagrange
>
> Lagrange: $0<\alpha_j^*$
>
> Hinge: $0<\alpha_j^*<C$

# 非线性SVM

## 支持向量

■ 支持向量：$\alpha_i^* > 0$ .

■ KKT对偶互补条件：$\boxed{\alpha_i^* g_i(\boldsymbol{w}^*) = 0}$ $\boxed{\eta_i^* \xi_i = 0.}$ $\boxed{\frac{\partial}{\partial \xi_i} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = C - \alpha_i - \eta_i = 0.}$

$\alpha_i^* \left[ y^i \left( (\boldsymbol{w}^*)^T \boldsymbol{x}^i + b \right) - 1 + \xi_i \right] = 0,\ \eta_i^* \xi_i = 0.$

■ 如果 $\alpha_i^* = 0$ ，$y^i \left( (\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) \geq 1$

■ 如果 $0 < \alpha_i^* < C$ ，$y^i \left( (\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) = 1$

■ 如果 $\alpha_i^* = C$ ，$y^i \left( (\boldsymbol{w}^*)^T \boldsymbol{x}^i + b_i \right) \leq 1$

- $0 < \xi_i < 1$
- $\xi_i = 1$
- $\xi_i > 1$