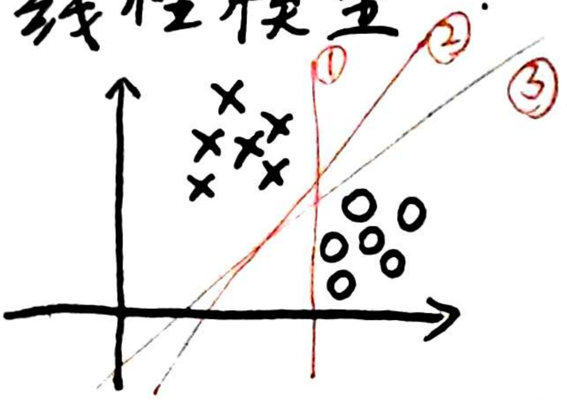
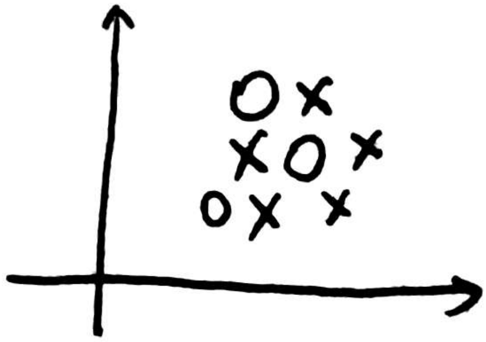
支持向量机 SVM

一. 线性模型

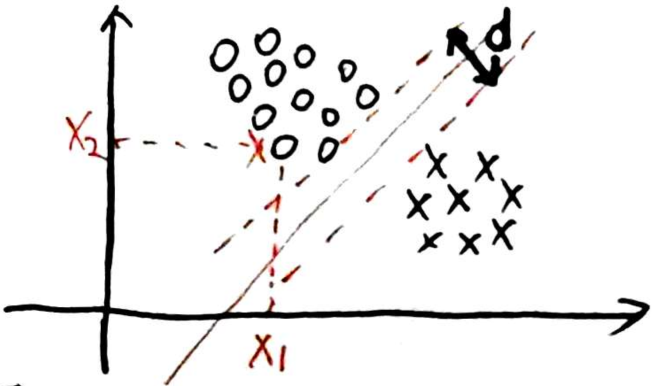




线性可分训练集　　非线性可分

可以证明：如果∃一条直线可以分开两个样本集，则有无数条直线可以将其分开

$X=\begin{bmatrix}x_1\\x_2\end{bmatrix}$



最大化间隔

- d：间隔
- 支持向量：平行线插到的向量 support Vector

定义：

① 训练数据及标签

$(x_1, y_1)\ (x_2, y_2) \cdots (x_n, y_n)$

向量 $x_1=\begin{bmatrix}x_{11}\\x_{12}\\\vdots\\x_{1n}\end{bmatrix}$　标签 $y_i=+1/-1$

② 线性模型：$(\underline{w}, \underline{b})$　$\underline{w^T x + b = 0}$ 超平面

向量 $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$　常数

③ 一个训练集线性可分是指：

$\{(x_i, y_i)\}_{i=1\sim N}$　$\exists (w, b)$ 使 $\forall i = 1 \sim N$ 有：

① 若 $y_i = +1$，则 $w^T x_i + b \geq 0$

② 若 $y_i = -1$，则 $w^T x_i + b < 0$

即 $y_i(w^T x_i + b) \geq 0$

SVM 优化问题：（凸优化）

$$\boxed{\begin{aligned} &\min \quad \frac{1}{2}\|w\|^2 \\ &s.t. \quad y_i[w^T x_i + b] \geq 1 \quad , i = 1 \sim N \end{aligned}}$$

事实1：$w^T x + b = 0$ 与 $a w^T x + ab = 0$ 是同一个平面.

$a \in R^+$

事实2：点到平面距离　平面：$w_1 x + w_2 y + b = 0$

点：(x_0, y_0)

$$d = \frac{|w_1 x_0 + w_2 y_0 + b|}{\sqrt{w_1^2 + w_2^2}}$$

x_0 到 $w^T x + b = 0$

$$d = \frac{|w^T x_0 + b|}{\|w\|}$$

可以用 a 缩放 $(w, b) \rightarrow (aw, ab)$

使在支持向量 x_0 上，有 $|w^T x_0 + b| = 1$

此时支持向量与平面距离 $d = \frac{1}{\|w\|}$

二. 非线性模型

事先设定的参数　　正则项

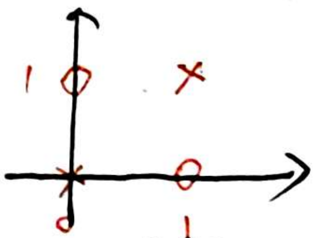
① 最小化：$\frac{1}{2}\|w\|^2 + \boxed{C\sum_{i=1}^{N}\varepsilon_i}$ 松弛变量

s.t. ：① $y_i[w^T\varphi(x_i)+b] \geq 1-\varepsilon_i \quad i=1\sim N$

② $\varepsilon_i \geq 0$

② 高维映射 $\varphi(x)$

$$\underset{\text{低维}}{X} \xrightarrow{\varphi} \underset{\text{高维}}{\varphi(X)}$$



异或问题

$x_1=\begin{bmatrix}0\\0\end{bmatrix}\in C_1 \quad x_2=\begin{bmatrix}1\\1\end{bmatrix}\in C_1$

$x_3=\begin{bmatrix}1\\0\end{bmatrix}\in C_2 \quad x_4=\begin{bmatrix}0\\1\end{bmatrix}\in C_2$

$\varphi(x)\ x=\begin{bmatrix}a\\b\end{bmatrix} \xrightarrow{\varphi} \varphi(x)=\begin{bmatrix}a^2\\b^2\\a\\b\\ab\end{bmatrix}$

$w=\begin{bmatrix}-1\\-1\\-1\\-1\\6\end{bmatrix} \quad b=1$

$\underset{1}{\varphi(x_1)}=\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix} \quad \underset{3}{\varphi(x_2)}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix} \in C_1$

$\underset{-1}{\varphi(x_3)}=\begin{bmatrix}1\\0\\1\\0\\0\end{bmatrix} \quad \underset{-1}{\varphi(x_4)}=\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix} \in C_2$

维度越高，被线性分开的概率越大.

$\varphi(x)$ 是无限维的.

我们可以不知道无限维映射 $\varphi(x)$ 的显示表达，我们只要知道一个核函数

一个数 $\underline{K(X_1, X_2)} = \varphi(X_1)^T \varphi(X_2)$ ← $\varphi(X_1)$ 与 $\varphi(X_2)$ 两个无限维向量的内积

则①这个优化式仍然可解.

核函数：高斯核 $K(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|^2}{2\delta^2}}$

多项式核 $K(x_1, x_2) = (x_1^T x_2 + 1)^{\underline{d}}$ ← 多项式阶数

$K(X_1, X_2)$ 能写成 $\varphi(x_1)^T \varphi(x_2)$ 的充要条件

① $K(x_1, x_2) = K(x_2, x_1)$　交换性

② $\forall \underline{\underline{c_i}}, \underline{\underline{x_i}}$ $(i = 1 \sim N)$ 有：

常数　向量　$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j K(x_i, x_j) \geq 0$.　半正定性

优化理论

原问题：

$$\boxed{\begin{aligned} &\min : f(w) \\ &s.t. : g_i(w) \leq 0, \ i = 1 \sim k \\ &\qquad\ \ h_i(w) = 0, \ i = 1 \sim m \end{aligned}}$$

对偶问题：

① 定义：$L(\underset{K}{w},\alpha,\underset{M}{\beta}) = f(w) + \sum_{i=1}^{K}\alpha_i g_i(w) + \sum_{i=1}^{M}\beta_i h_i(w)$

$= f(w) + \alpha^T g(w) + \beta^T h(w)$

② $$\boxed{\begin{aligned} &\max:\ \theta(\alpha,\beta) = \overset{\text{求最小值}}{\underset{\text{所有}w}{\inf}}\{L(w,\alpha,\beta)\} \\ &\text{s.t.}\quad \alpha_i \geq 0\ (i=1\sim k) \end{aligned}}$$ α,β确定

定理：如果w^*是原问题的解，而α^*,β^*是对偶问题的解，则有 $f(w^*) \geq \theta(\alpha^*,\beta^*)$

证明：$\theta(\alpha^*,\beta^*) = \inf_{\forall w}\{L(w,\alpha^*,\beta^*)\}$

$\leq L(w^*,\alpha^*,\beta^*)$

$= f(w^*) + \sum_{i=1}^{K}\underbrace{\alpha_i^*}_{\geq 0}\underbrace{g_i(w^*)}_{\leq 0} + \sum_{i=1}^{M}\beta_i^*\underbrace{h_i(w^*)}_{=0}$

$\leq f(w^*)$

#

定义：$G = f(w^*) - \theta(\alpha^*,\beta^*) \geq 0$

G叫做原问题与对偶问题的间距

对于某些特定的优化问题，可以证明 $G=0$.

强对偶定理：若$f(w)$是凸函数，且$g(w) = Aw+b$，$h(w) = cw+d$，则此优化问题的原问题与对偶问题间距为0. 即 $f(w^*) = \theta(\alpha^*,\beta^*)$

对于$\forall i=1\sim k$ 或者$\alpha_i^*=0$ 或者$g_i(w^*)=0$　KKT条件

原问题:

$$\min: \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i$$

$$\text{s.t.}: \ ① \ y_i\left[w^T\varphi(x_i)+b\right] \geq 1-\varepsilon_i$$

$$② \ \varepsilon_i \geq 0$$

$\Downarrow$

$$\min: \ \frac{1}{2}\|w\|^2 - C\sum_{i=1}^{N}\varepsilon_i$$

$$\text{s.t.}: \ ① \ 1+\varepsilon_i - y_i w^T\varphi(x) - y_i b \leq 0$$

$$② \ \varepsilon_i \leq 0$$

对偶问题:

$$\max: \ \theta(\alpha,\beta) = \inf_{(w,\varepsilon_i,b)}\left\{\frac{1}{2}\|w\|^2 \cancel{- C\sum_{i=1}^{N}\varepsilon_i} \cancel{+ \sum_{i=1}^{N}\beta_i\varepsilon_i} + \sum_{i=1}^{N}\alpha_i\left[1+\cancel{\varepsilon_i} - y_i w^T\varphi(x) \cancel{- y_i b}\right]\right\}$$

$$\text{s.t.}: \ \alpha_i \geq 0, \quad i=1\sim N$$

$$\beta_i \geq 0$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w - \sum_{i=1}^{N}\alpha_i y_i \varphi(x_i) = 0$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow -C + \beta_i + \alpha_i = 0$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow -\sum_{i=1}^{N}\alpha_i y_i = 0$$

代入 $\theta(\alpha,\beta)$

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

$$\begin{aligned}\frac{1}{2}\|w\|^2&=\frac{1}{2}w^Tw\\&=\frac{1}{2}\Big(\sum_{i=1}^{N}\alpha_i y_i\varphi(x_i)\Big)^T\Big(\sum_{j=1}^{N}\alpha_j y_j\varphi(x_j)\Big)\\&=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\underline{\varphi(x_i)^T\varphi(x_j)}=k(x_i,x_j)\end{aligned}$$

$$\begin{aligned}-\sum_{i=1}^{N}\alpha_i y_i w^T\varphi(x_i)&=-\sum_{i=1}^{N}\alpha_i y_i\Big(\sum_{j=1}^{N}\alpha_j y_j\varphi(x_j)\Big)^T\varphi(x_i)\\&=-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\underline{\varphi(x_j)^T\varphi(x_i)}=k(x_i,x_j)\end{aligned}$$

~~对偶问题:~~

$$\boxed{\begin{aligned}&\max:\ \theta(\alpha,\cancel{\beta})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(x_i,x_j)\\&\text{s.t.}\ \ ①\ 0\le\alpha_i\le C\\&\qquad ②\ \sum_{i=1}^{N}\alpha_i y_i=0\end{aligned}}$$

测试流程:

测试样本 X,

$$\begin{cases}\text{若}\ \ w^T\varphi(x)+b\ge 0,\ \text{则}\ y=+1\\ \text{若}\ \ w^T\varphi(x)+b<0,\ \text{则}\ y=-1\end{cases}$$

$$\begin{aligned}w^T\varphi(x)&=\sum_{i=1}^{N}[\alpha_i y_i\varphi(x_i)]^T\varphi(x)\\&=\sum_{i=1}^{N}\alpha_i y_i\varphi(x_i)^T\varphi(x)\\&=\sum_{i=1}^{N}\alpha_i y_i k(x_i,x)\end{aligned}$$

KKT条件：$\forall i=1\sim N$

① 要么 $\beta_i=0$，要么 $\varepsilon_i=0$

② 要么 $\alpha_i=0$，要么 $1+\varepsilon_i-y_i w^T\varphi(x_i)-y_i b=0$

取一个 $0<\alpha_i<C \Rightarrow \beta_i=C-\alpha_i>0$

此时 $\beta_i\neq 0 \Rightarrow \varepsilon_i=0$

$\alpha_i\neq 0 \Rightarrow 1-y_i w^T\varphi(x_i)-y_i b=0$

$$b=\frac{1-y_i w^T\varphi(x_i)}{y_i}=\frac{1-y_i\sum_{j=1}^{N}\alpha_j y_j K(x_j,x_i)}{y_i}$$

SVM流程

1. 训练流程，求 α

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

s.t. ① $0\le\alpha_i\le C$

② $\sum_{i=1}^{N}\alpha_i y_i=0$

求 b，完成训练，见上

2. 测试流程，考察测试数据X，预测类别Y

←