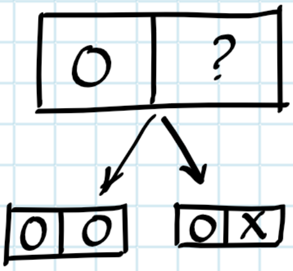
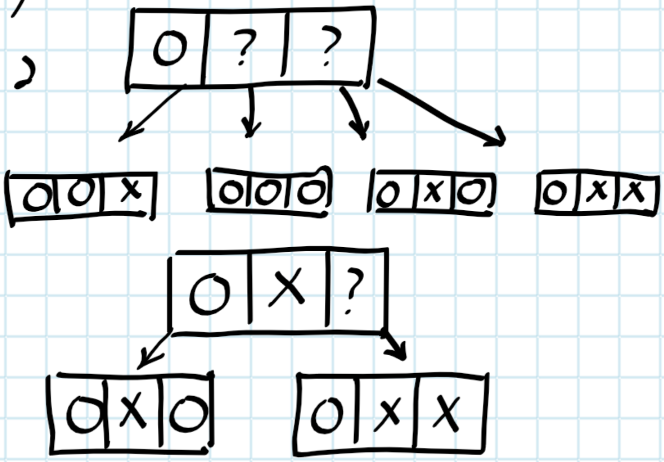
没有免费午餐定理 (No Free Lunch Theorem)

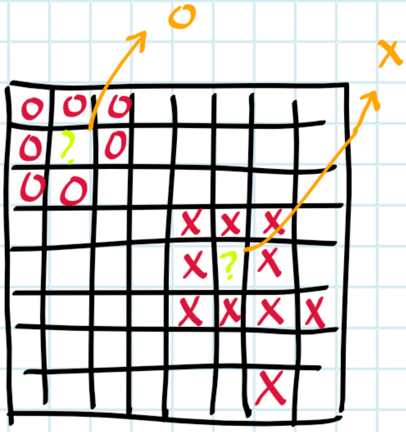
↳ 如果我们不对特征空间有先验假设，则所有算法的平均表现是一样的！



O: Class 1

X: Class 2





O 花瓣

X 蜜蜂

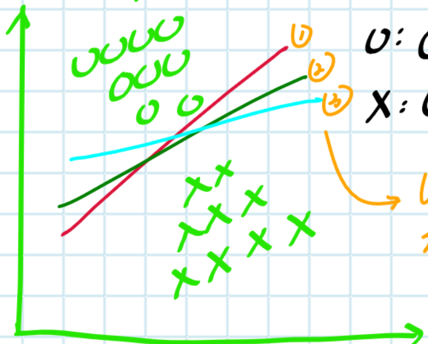
特征差距小的样本更有可能是同一类～

Vapnik

支持向量机 (Support Vector Machine) → 小样本方法

↳ 小样本依旧能获得很好的结果!!!

(1) 线性模型

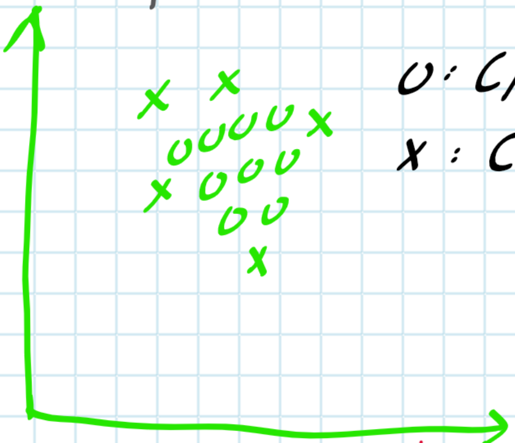


O: Class 1

X: Class 2

Which one is the best?

线性可分 (Linear Separable)



O: Class 1

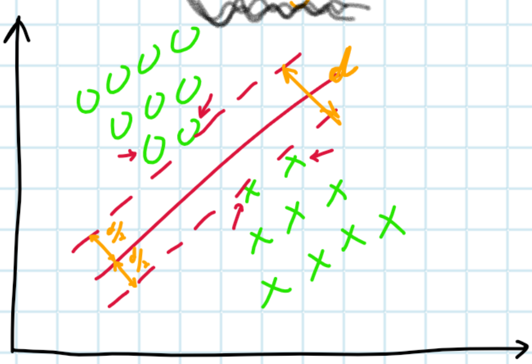
X: Class 2

非线性可分 (None Linear Separable)

performance measure

性能指标

Why?



d: Margin

支持向量 (Support Vectors)

"→" 所指的数据

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 标签 $(y=+1/-1)$

② 线性模型 (W,b) $W^Tx+b=0$ (超平面) (Hyperplane)

向量 常数 常数

$X=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$ $w=\begin{bmatrix}w_1\\w_2\\\vdots\\w_n\end{bmatrix}$

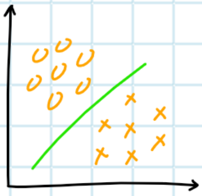
③ 一个训练集线性可分是指:

$\{(x_i,y_i)\}_{i=1\sim N}$

$\exists(w,b)$, 使: 对 $\forall i=1\sim N$, 有:

(a) 若 $y_i=+1$, 则 $w^Tx_i+b>0$

(b) 若 $y_i=-1$, 则 $w^Tx_i+b<0$



$y_i[w^Tx_i+b]\geq 0$ ---- 公式1

优化问题: (凸优化问题, 二次规划问题)

最小化 (Minimize): $\|w\|^2\cdot\frac{1}{2}$ 求导方便

限制条件 (subject to): $y_i(w^Tx_i+b)\geq 1, i=1\sim N$

事实1: $w^Tx+b=0$ 与 $aw^Tx+ab=0$ 是同一个平面 $a\in R^+$

若 (w,b) 满足公式1, 则 (aw,ab) 也满足公式1

事实2: 点到面的距离公式:

平面: $w_1x+w_2y+b=0$, 则点 (x_0,y_0) 到此面的距离

$$d=\frac{|w_1x_0+w_2y_0+b|}{\sqrt{w_1^2+w_2^2}}$$

向量 X_0 到超平面 $w^Tx+b=0$ 的距离

$$d=\frac{|w^Tx_0+b|}{\|w\|}$$

$\|w\| = \sqrt{w_1^2+w_2^2+\cdots+w_n^2}$

我们可以用a去缩放

$(w,b)\rightarrow(aw,ab)$

最终使在支持向量 x_0 上有:

$|w^Tx_0+b|=1$

此时支持向量与平面距离:

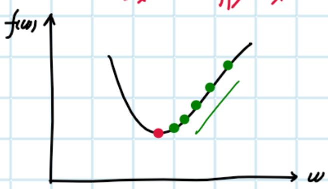
$d=1/\|w\|$

二次规划 (Quadratic Programming):

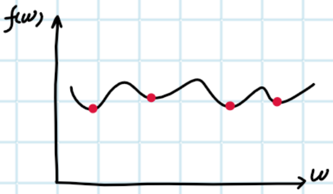
① 目标函数 (Objective Function) 是二次项

② 限制条件是一次项

要么无解/要么只有一个极值



⇒ 全局最优解



⇒ Hard to Solve

① SVM是最大化间隔(Margin)的分类算法

② 优化问题

训练样本 $\{(x_i,y_i)\}_{i=1\sim N}$ 向量 标签

③ 最小化: $\frac{1}{2}\|w\|^2$

限制条件: $y_i[w^Tx_i+b]\geq 1$ $(i=1\sim N)$

SVM处理非线性:

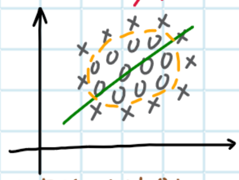
① 最小化: $\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i$ 松弛变量 (Slack Variable)

② 限制条件: $y_i[w^Tx_i+b]\geq 1-\xi_i$ $(i=1\sim N)$

$\xi_i\geq 0$

$C\sum_{i=1}^{N}\xi_i$: 正则项 (Regulation Term)

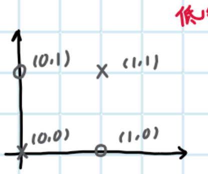
C 事先设定的参数



定义一个高维的映射 $\varphi(x)$:

$x\xrightarrow{\varphi}\varphi(x)$

低维 高维



⇒ 异或问题

$x_1=\begin{bmatrix}0\\0\end{bmatrix}\in C_1$ $x_2=\begin{bmatrix}1\\1\end{bmatrix}\in C_1$ $x_3=\begin{bmatrix}1\\0\end{bmatrix}\in C_2$ $x_4=\begin{bmatrix}0\\1\end{bmatrix}\in C_2$

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

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

$\varphi(x_3)=\begin{bmatrix}1\\0\\1\\0\\0\end{bmatrix}$ $\varphi(x_4)=\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}$ $\in C$

我们可以不知道无限维映射 $\varphi(x)$ 的显式表达, 只需要知道一个核函数 (Kernel Function)

$K(x_1,x_2)=\varphi(x_1)^T\varphi(x_2)$

则 ①这个优化式依然可解

核函数

① $K(x_1,x_2)=e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}$ (高斯核)

$=\varphi(x_1)^T\varphi(x_2)$

② $K(x_1,x_2)=(x_1^Tx_2+1)^d$ (多项式核)

$=\varphi(x_1)^T\varphi(x_2)$

$K(x_1,x_2)\Rightarrow\varphi(x_1)^T\varphi(x_2)$ 的充要条件:

① $K(x_1,x_2)=K(x_2,x_1)$ (交换性)

② $\forall c_i, x_i$ $(i=1\sim N)$, 有: (半正定性)

$\sum_{i=1}^{N}\sum_{j=1}^{N}c_ic_jK(x_i,x_j)\geq 0$

优化理论:

① 《Convex optimization》

② 《Non linear Programming》

原问题 (Prime Problem) → 非常普遍

最小化: $f(w)$

限制条件: $g_i(w) \le 0 \quad (i=1\sim K)$

$h_i(w) = 0 \quad (i=1\sim M)$

对偶问题 (Dual Problem):

① 定义: $L(w,\alpha,\beta)$

$$= f(w) + \sum_{i=1}^{K}\alpha_i g_i(w) + \sum_{i=1}^{M}\beta_i h_i(w)$$

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

$$g(w) = \begin{bmatrix} g_1(w) \\ g_2(w) \\ \vdots \\ g_K(w) \end{bmatrix} \qquad h(w) = \begin{bmatrix} h_1(w) \\ h_2(w) \\ \vdots \\ h_M(w) \end{bmatrix}$$

② 对偶问题的定义:

最大化: $\theta(\alpha,\beta) = \inf_{\text{所有} w}\{L(w,\alpha,\beta)\}$ 求最小值

限制条件: $\alpha_i \ge 0 \quad (i=1\sim K)$

定理: 如果 w^* 是原问题的解, 而 α^*, β^* 是对偶问题的解, 则有:

$$f(w^*) \ge \theta(\alpha^*, \beta^*)$$

证: $\theta(\alpha^*,\beta^*) = \inf_{\text{所有} w}\{L(w,\alpha^*,\beta^*)\} \le L(w^*,\alpha^*,\beta^*)$

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

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

G 叫做原问题与对偶问题的间距 (Duality Gap).

对于某些特定的优化问题, 可以证明: $G = 0$

强对偶定理:

若 $f(w)$ 为凸函数, 且 $g(w) = Aw+b$, $h(w) = Cw+b$, 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)

将支持向量机的原问题转化为对偶问题

最小化 $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

限制条件 ① $y_i[w^T\varphi(x_i)+b] \ge 1-\xi_i$

② $\xi_i \ge 0$

→

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

① $y_i[w^T\varphi(x_i)+b] \le 1+\xi_i$

② $\xi_i \le 0$

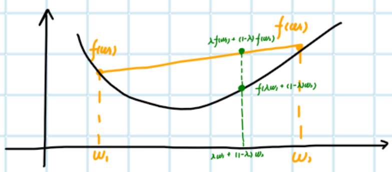
↓

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

① $1+\xi_i - y_i w^T\varphi(x_i) - y_i b \le 0$

② $\xi_i \le 0$

关于凸函数:



$\forall w_1, w_2, \ \forall \lambda \in [0,1]$,

$f(\lambda w_1 + (1-\lambda)w_2) \le \lambda f(w_1) + (1-\lambda)f(w_2)$ } 数学定义

对偶问题:

最大化 $\theta(\alpha,\beta) = \inf_{\text{所有}(w_i,\xi_i,b)}\left\{\frac{1}{2}\|w\|^2 - C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i[1+\xi_i - y_i w^T\varphi(x_i) - y_i b]\right\}$

限制条件 $\alpha_i \ge 0 \quad \beta_i \ge 0 \quad (i=1\sim N)$

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w - \sum_{i=1}^{N}\alpha_i y_i \varphi(x_i) = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow -C+\beta_i+\alpha_i = 0 \\ \frac{\partial L}{b} = 0 \Rightarrow -\sum_{i=1}^{N}\alpha_i y_i = 0 \end{cases}$$

$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$ $\quad f(w) \to$ 数

$\frac{\partial f}{\partial w} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_m} \end{bmatrix}$

若 $f(w) = \frac{1}{2}\|w\|^2$

则 $\frac{\partial f}{\partial w} = w$

若 $f(w) = w^T X$

则 $\frac{\partial f}{\partial w} = X$

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

限制条件: ① $0 \le \alpha_i \le C$

② $\sum_{i=1}^{N}\alpha_i y_i = 0$

凸优化问题 → SMO 算法

测试流程:

测试样本 X,

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

SVM 算法:

① 训练流程

输入训练样本 $\{(x_i,y_i)\}_{i=1\sim N}$

(解优化问题)

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

限制条件: ① $0 \le \alpha_i \le C$

② $\sum_{i=1}^{N}\alpha_i y_i = 0$

算 b, 找一个 $0 < \alpha_i < C$,

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

② 测试流程

输入测试样本 X

$$\begin{cases} \text{若 } \sum_{i=1}^{N}\alpha_i y_i K(x_i,X)+b \ge 0, \text{ 则 } y=+1 \\ \text{若 } \sum_{i=1}^{N}\alpha_i y_i K(x_i,X)+b < 0, \text{ 则 } y=-1 \end{cases}$$

SVM内核函数:

Linear(线性内核): $K(x,y) = x^T y$

Poly(多项式核): $K(x,y) = (x^T y + 1)^d$

Rdf(高斯径向基函数核): $K(x,y) = e^{\frac{\|x-y\|^2}{\sigma^2}}$

Tanh(Tanh核): $K(x,y) = \tanh(\beta x^T y + b)$ $\quad \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$