§8.1对偶SVM动机

非线性支撑向量机模型

最佳的(w,b)=?

$$\min_{w} \frac{1}{2}w^Tw$$
Subject to $y_n(w^Tz_n+b)\ge 1$, $z_n=\Phi(x_n)$, for $n=1,\dots,N$

① $Q=\begin{bmatrix}0 & 0_d^T\\ 0_d & I_d\end{bmatrix}$, $p=0_{d+1}$, $a_n^T=y_n[1\ x_n^T]$, $c_n=1$

② $\begin{bmatrix}b\\ w\end{bmatrix}\leftarrow QP(Q,p,A,c)$

③ 返回最终的b和w作为学到的g_{SVM}

如果 $\tilde d$ 很大，甚至无穷大，挑战巨大

QP针对$(\tilde d+1)$个变量和N个约束条件求解

目的：SVM算法求解可以不依赖于$\tilde d$吗?

∅用Lagrange乘子α_n构造Lagrange函数

$$L(b,w,\alpha)=\frac{1}{2}w^Tw+\sum_{n=1}\alpha_n(1-y_n(w^Tz_n+b))$$

SVM优化问题转化为

$$\min_{b,w}\left(\max_{all\ \alpha_n\ge 0}L(b,w,\alpha)\right)=\min_{b,w}\left(\infty \text{ if violate},\ \frac{1}{2}w^Tw \text{ if feasible}\right)=\frac{1}{2}w^Tw$$

violate (b,w): $\max_{all\ \alpha_n\ge0}(\frac{1}{2}w^Tw+\sum\alpha_n\cdot$正数$)\to\infty$

feasible (b,w): $\max_{all\ \alpha_n\ge0}(\frac{1}{2}w^Tw+\sum\alpha_n$非正数$)=\frac{1}{2}w^Tw$

§8.2对偶SVM拉格朗日对偶

在所有的$\alpha_n\ge 0$的中挑选任意一个α' ($\because$ max $\ge$ any)

$$\min_{b,w}\left(\max_{all\ \alpha_n\ge0}\mathcal{L}(b,w,\alpha)\right)\ge\min_{b,w}(\mathcal{L}(b,w,\alpha'))$$

如果$\alpha'\ge0$是上式右边 $\max_{all\ \alpha_n\ge0}\mathcal{L}(b,w,\alpha)$中的最佳值 ($\because$ best is one of any)

$$\underbrace{\min_{b,w}\left(\max_{all\ \alpha_n\ge0}\mathcal{L}(b,w,\alpha)\right)}_{\text{等价于原SVM问题}}\ge\underbrace{\max_{all\ \alpha_n\ge0}\left(\min_{b,w}(\mathcal{L}(b,w,\alpha'))\right)}_{\text{Lagrange Dual Problem}}$$

弱对偶

原问题(求解b,w)与拉格朗日对偶问题(求解α)的关系

若满足 { 原问题凸函数; 原问题存在可行解; 约束条件线性性(非线性变换$\Phi(x)$) } = 二次规划QP问题 ⇒ 可取等构成强对偶问题

$$\min_{b,w}\left(\max_{all\ \alpha_n\ge0}\mathcal{L}(b,w,\alpha)\right)=\max_{all\ \alpha_n\ge0}\left(\min_{b,w}(\mathcal{L}(b,w,\alpha))\right)$$

equivto original SVM Primal Problem; Lagrange Dual Problem

等式两边对原问题求解和对对偶问题求解都能得到最优(b,w,α)

∅对偶问题求解

$$\max_{all\ \alpha_n\ge0}\left(\underbrace{\min_{b,w}\left(\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n(1-y_n(w^Tz_n+b))\right)}_{\mathcal{L}(b,w,\alpha)}\right)$$

• "括号"内的问题(inner problem)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,w,\alpha)}{\partial b}=-\sum_{n=1}^N\alpha_ny_n=0$$

$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0}\left(\min_{b,w}\left(\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n(1-y_n(w^Tz_n))-\sum_{n=1}^N\alpha_ny_nb\right)\right)$$

$$\max_{all\ \alpha_n\ge0}\left(\underbrace{\min_{b,w}\left(\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n(1-y_n(w^Tz_n+b))\right)}_{\mathcal{L}(b,w,\alpha)}\right)$$

• "括号"内的问题(inner problem)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,w,\alpha)}{\partial b}=-\sum_{n=1}^N\alpha_ny_n=0$$

① "括号"内的问题b取最佳解时：
$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0}\left(\min_{b,w}\left(\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n(1-y_n(w^Tz_n))\right)\right)$$

• "括号"内的问题(inner problem)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,w,\alpha)}{\partial w_i}=w_i-\sum_{n=1}^N\alpha_ny_nz_{n,i}=0 \Rightarrow w=\sum_{n=1}^N\alpha_ny_nz_n$$

② "括号"内的问题w取最佳解时：
$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0,w=\sum_{n=1}^N\alpha_ny_nz_n}\left(\min_{b,w}\left(\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n-w^Tw\right)\right)$$

"括号"内的问题w取最佳解时：
$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0,w=\sum_{n=1}^N\alpha_ny_nz_n}\left(\min_{b,w}\left(-\frac{1}{2}w^Tw+\sum_{n=1}^N\alpha_n\right)\right)$$

"括号"内的问题w取最佳解时：
$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0,w=\sum_{n=1}^N\alpha_ny_nz_n}\left(-\frac{1}{2}\left\|\sum_{n=1}^N\alpha_ny_nz_n\right\|^2+\sum_{n=1}^N\alpha_n\right)$$

如果(b,w,α)是原问题-对偶问题的最佳解(primal-dual optimal)：
- 原问题可行解(primal feasible)：$y_n(w^Tz_n+b)\ge1$ $1-y_n(w^Tz_n+b)\le 0$
- 对偶问题可行解(dual feasible)：$\alpha_n\ge0$ $\alpha_n\ge0$
- 对偶"括号"内优化解(dual-inner optimal)：$\sum_{n=1}^N\alpha_ny_n=0, w=\sum_{n=1}^N\alpha_ny_nz_n$ $\nabla L=0$
- 原问题"括号"内优化解(primal-inner optimal)：$\alpha_n(1-y_n(w^Tz_n+b))=0$

----称为KKT条件，为优化的充要条件   利用对偶问题求解最佳α后，再用KKT条件得到(b,w)

KKT条件

KKT条件将Lagrange乘子法所处理涉及等式的约束优化问题推广至不等式
1) 是拉格朗日取得可行解的必要条件；
2) 这就是以上分析的一个比较有意思的约束，称作松弛互补条件；
3)、4) 初始约束条件；
5) 不等式约束的 Lagrange Multiplier 需满足的条件

$$\begin{cases}\alpha_i\geqslant 0;\\ y_if(x_i)-1\geqslant0;\\ \alpha_i(y_if(x_i)-1)=0.\end{cases}$$

$\min f(x), s.t. h_i(x)=0, i=1,2,\dots,m; g_j(x)\le0, j=1,2,\dots,n$

$$L(x,\alpha,\beta)=f(x)+\sum_{i=1}^m\alpha_ih_i(x)+\sum_{j=1}^n\beta_jg_j(x)$$

$$\begin{aligned}\nabla_xL(x,\alpha,\beta)&=0 &(1)\\ \beta_jg_j(x)&=0,\ j=1,2,\dots,n &(2)\\ h_i(x)&=0,\ i=1,2,\dots,m &(3)\\ g_j(x)&\le0,\ j=1,2,\dots,n &(4)\\ \beta_j&\ge0,\ j=1,2,\dots,n &(5)\end{aligned}$$

§8.3求解SVM最佳值

支撑向量机的对偶问题求解

$$\max_{all\ \alpha_n\ge0,\sum_{n=1}^N\alpha_ny_n=0,w=\sum_{n=1}^N\alpha_ny_nz_n}\left(-\frac{1}{2}\left\|\sum_{n=1}^N\alpha_ny_nz_n\right\|^2+\sum_{n=1}^N\alpha_n\right)$$

标准的硬间隔SVM的对偶问题----求解最佳α：

$$\min_\alpha \frac{1}{2}\sum_{n=1}^N\sum_{m=1}^N\alpha_n\alpha_my_ny_mz_n^Tz_m-\sum_{n=1}^N\alpha_n$$
Subject to $\sum_{n=1}^N\alpha_ny_n=0$
$\alpha_n\ge0$, for $n=1,2,\dots,N$

- α的目标函数是二次函数、凸函数！N个变量
- α的约束条件是线性函数！$N+1$个约束条件

----二次规划(QP)问题！

QP有成熟方便的办法求优化解！

SVM对偶问题的一般求解：

最佳的α=?
$$\min_\alpha \frac{1}{2}\sum_{n=1}^N\sum_{m=1}^N\alpha_n\alpha_my_ny_mz_n^Tz_m-\sum_{n=1}^N\alpha_n$$
Subject to $\sum_{n=1}^N\alpha_ny_n=0$, $\alpha_n\ge0$, for $n=1,2,\dots,N$

$u=[\alpha_1\ \cdots\ \alpha_n\ \cdots\ \alpha_N]^T$, $p=-1_N$

$$Q=\begin{bmatrix}y_1y_1z_1^Tz_1 & \cdots & y_1y_Nz_1^Tz_N\\ \vdots & y_ny_mz_n^Tz_m & \vdots\\ y_Ny_1z_N^Tz_1 & \cdots & y_Ny_Nz_N^Tz_N\end{bmatrix}$$

二次规划(QP)的求解：

最佳的$u\leftarrow QP(Q,p,A,c)$
$$\min_u \frac{1}{2}u^TQu+p^Tu$$
Subject to $a_m^Tu\ge c_m$, for $m=1,2,\dots,M$

$a_\ge^T=[y_1\ \cdots\ y_n\ \cdots\ y_N]$, $c_\ge=0$
$a_\le^T=[-y_1\ \cdots\ -y_n\ \cdots\ -y_N]$, $c_\le=0$
$a_1^T=[1\ \cdots\ 0\ \cdots\ 0]$,
$a_n^T=[0\ \cdots\ 1\ \cdots\ 0]$, $c_n=0$, $M=N$
$a_N^T=[0\ \cdots\ 0\ \cdots\ 1]$,

利用二次规划(QP)实现支撑向量机对偶问题求解

① $Q=\begin{bmatrix}y_1y_1z_1^Tz_1 & \cdots & y_1y_Nz_1^Tz_N\\ \vdots & y_ny_mz_n^Tz_m & \vdots\\ y_Ny_1z_N^Tz_1 & \cdots & y_Ny_Nz_N^Tz_N\end{bmatrix}$, $p=-1_N$,

$a_n^T=[0\ \cdots\ 1\ \cdots\ 0]$, $c_n=0$,
$a_\ge^T=[y_1\ \cdots\ y_n\ \cdots\ y_N]$, $c_\ge=0$,
$a_\le^T=[-y_1\ \cdots\ -y_n\ \cdots\ -y_N]$, $c_\le=0$,

② $[\alpha_1\ \cdots\ \alpha_n\ \cdots\ \alpha_N]^T\leftarrow QP(Q,p,A,c)$

③ 返回最终的α

求解最佳的(b,w)

如果(b,w,α)是原问题-对偶问题的最佳解(primal-dual optimal)：
- 原问题可行解(primal feasible)：$y_n(w^Tz_n+b)\ge1$
- 对偶问题可行解(dual feasible)：$\alpha_n\ge0$
- 对偶"括号"内优化解(dual-inner optimal)：$\sum_{n=1}^N\alpha_ny_n=0, w=\sum_{n=1}^N\alpha_ny_nz_n$
- 原问题"括号"内优化解(primal-inner optimal)：$\alpha_n(1-y_n(w^Tz_n+b))=0$

----称为KKT条件，为优化的充要条件

挑选任意一个 $\alpha_n>0$ 的样本，$b=y_n-w^Tz_n$ $\alpha_n>0$ 的样本是支撑向量

§8.4支撑向量讨论

SVM原问题 { 分类面由边界上样本确定(其他样本不起作用); 边界上样本被称为支撑向量(候选) }

SVM对偶问题 { $\alpha_n>0$ 样本落在分类面边界上 ⇒ SV($\alpha_n>0$) ⊆ SV候选(边界上样本); $\alpha_n>0$ 样本(z_n,y_n)被称为支撑向量(候选) }

→ $w=\sum_{n=1}^N\alpha_ny_nz_n=\sum_{SV}\alpha_ny_nz_n$ (支撑向量)

$b=y_n-w^Tz_n$ (任一支撑向量(z_n,y_n))

SVM：
$$w_{SVM}=\sum_{n=1}^N\alpha_n(y_nz_n)$$
α_n由对偶问题的解确定

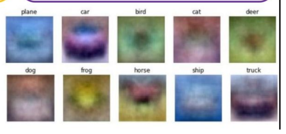
PLA：
$$w_{PLA}=\sum_{n=1}^N\beta_n(y_nz_n)$$
β_n由分错的样本确定

w 是 y_nz_n 的线性组合

w 被数据 y_nz_n 所表达

w 体现出"模式"

SVM： 仅通过SV表达w



SVM的原问题求解：
$$\min_w \frac{1}{2}w^Tw$$
Subject to $y_n(w^Tz_n+b)\ge1$, for $n=1,2,\dots,N$
- $(\tilde d+1)$ 个变量和N 个约束条件
- 求解最佳(b,w)

SVM的对偶问题求解：
$$\min_\alpha \frac{1}{2}\sum_{n=1}^N\sum_{m=1}^N\alpha_n\alpha_my_ny_mz_n^Tz_m-\sum_{n=1}^N\alpha_n$$
Subject to $\sum_{n=1}^N\alpha_ny_n=0$, $\alpha_n\ge0$, for $n=1,2,\dots,N$
- N 个变量和 $N+1$ 个约束条件
- 求解最佳α，确定支撑向量

两种方法都能得到最佳解(b,w)获得最大间隔分类面$g_{SVM}=\mathrm{sign}(w^T\phi(x_n)+b)$

于是，对任意训练样本 (x_i,y_i)，总有 $\alpha_i=0$ 或 $y_if(x_i)=1$。若 $\alpha_i=0$，则该样本将不会在式(6.12) 的求和中出现，也就不会对 $f(x)$ 有任何影响；若 $\alpha_i>0$，则必有 $y_if(x_i)=1$，所对应的样本点位于最大间隔边界上，是一个支持向量。这显示出支持向量机的一个重要性质：训练完成后，大部分的训练样本都不需保留，最终模型仅与支持向量有关。

§8.4核函数SVM

$$\min_\alpha \frac{1}{2}\sum_{n=1}^N\sum_{m=1}^N\alpha_n\alpha_my_ny_mz_n^Tz_m-\sum_{n=1}^N\alpha_n$$
Subject to $\sum_{n=1}^N\alpha_ny_n=0$
$\alpha_n\ge0$, for $n=1,2,\dots,N$

$$\min_\alpha \frac{1}{2}\alpha^TQ\alpha-1^T\alpha$$
Subject to $y^T\alpha=0$
$\alpha_n\ge0$, for $n=1,2,\dots,N$

研究动机⇒不想依赖$\tilde d$

N个变量与$N+1$个约束条件

且对偶问题Q矩阵稠密

$$Q=\begin{bmatrix}y_1y_1z_1^Tz_1 & \cdots & y_1y_Nz_1^Tz_N\\ \vdots & y_ny_mz_n^Tz_m & \vdots\\ y_Ny_1z_N^Tz_1 & \cdots & y_Ny_Nz_N^Tz_N\end{bmatrix}$$

提高效率 $z_n^Tz_m=\Phi(x_n)^T\Phi(x_m)$

∅ 计算元素 $q_{n,m}=y_ny_mz_n^Tz_m$ 做内积运算，代价$O(\tilde d)$

Φ变换+内积运算

$\Rightarrow K_\Phi(x,x')=\Phi^T(x)\Phi(x')$

二次多项式$\Phi_2(x)$的快速内积计算

$\Phi_2(x)=(1,x_1,x_2,\dots,x_d,x_1^2,x_1x_2,\dots,x_1x_d,x_2x_1,x_2^2,\dots,x_2x_d,\dots,x_d^2)$

$$K_{\Phi_2}(x,x')=\Phi_2^T(x)\Phi_2(x')=1+\sum_{i=1}^dx_ix_i'+\sum_{i=1}^d\sum_{j=1}^dx_ix_jx_i'x_j'$$
$$=1+\sum_{i=1}^dx_ix_i'+\sum_{i=1}^dx_ix_i'\sum_{j=1}^dx_jx_j'$$
$$=1+x^Tx'+(x^Tx')(x^Tx')$$

计算复杂度从$O(\tilde d)$下降到$O(d)$

$q_{n,m}=y_ny_mz_n^Tz_m=y_ny_m\Phi^T(x_n)\Phi(x_m)=y_ny_mK_\Phi(x_n,x_m)$

$w=\sum_{n=1}^N\alpha_ny_nz_n=\sum_{SV}\alpha_ny_nz_n$

支撑SV(z_m,y_m)：$b=y_m-w^Tz_m=y_m-(\sum_{SV}\alpha_ny_nz_n)^Tz_m=y_m-\sum_{SV}\alpha_ny_n\Phi^T(x_n)\Phi(x_m)$ $K_\Phi(x_n,x_m)$ $O(\tilde d)\to O(d)$

输入测试样本 x：$g_{SVM}=\mathrm{sign}(w^T\Phi(x)+b)=\mathrm{sign}((\sum_{SV}\alpha_ny_n\Phi(x_n))^T\Phi(x)+b)=\mathrm{sign}(\sum_{SV}\alpha_ny_n\Phi(x_n)\Phi(x)+b)$ $K_\Phi(x_n,x)$

利用二次规划(QP)实现核函数支撑向量机求解

① $q_{n,m}=y_ny_mK_\Phi(x_n,x_m)$，$p=-1_N$，由约束条件得到$(A,c)$

② $\alpha\leftarrow QP(Q,p,A,c)$

③ 任选一支撑SV(x_m,y_m)：$b\leftarrow(y_m-\sum_{SV}\alpha_ny_nK_\Phi(x_n,x_m))$

④ 对于测试样本 x：$g_{SVM}=sign(\sum_{SV}\alpha_ny_nK_\Phi(x_n,x)+b)$

- 步骤①的时间复杂度：$O(N^2)\cdot$(kernel evaluation)
- 步骤②的开销 ：N个变量，$N+1$个约束
- 步骤③和④的复杂度：$O(\#SV)\cdot$(kernel evaluation)

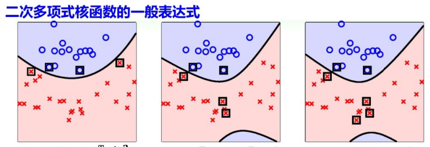
二次多项式核函数的一般表达式

$\Phi_2(x)$	$K_2(x,x')$
$(1,x_1,\dots,x_d,x_1^2,x_1x_2,\dots,x_d^2)$	$1+x^Tx'+(x^Tx')^2$
$(1,\sqrt{2}x_1,\dots,\sqrt{2}x_d,x_1^2,x_1x_2,\dots,x_d^2)$	$1+2x^Tx'+(x^Tx')^2$
$(1,\sqrt{2\gamma}x_1,\dots,\sqrt{2\gamma}x_d,x_1^2,x_1x_2,\dots,x_d^2)$	$1+2\gamma x^Tx'+\gamma^2(x^Tx')^2$

$$K_2(x,x')=(1+\gamma x^Tx')^2 \qquad \gamma>0$$

不同的二次多项式变换：
- 升维次数相同
- 内积结果不同 ➡ 分类面不同

二次多项式核函数的一般表达式



$(1+0.001x^Tx')^2$ $1+x^Tx'+(x^Tx')^2$ $(1+1000x^Tx')^2$

- 核函数不同 ➡ 支撑向量(SVs)不同，分类面(g_{SVM})不同
- 核函数变化 ➡ Margin也会变化

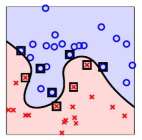
多项式核函数的一般表达式

$K_2(x,x')=(\zeta+\gamma x^Tx')^2$ $\gamma>0,\zeta\ge0$

$K_3(x,x')=(\zeta+\gamma x^Tx')^3$ $\gamma>0,\zeta\ge0$

$\vdots$

$K_Q(x,x')=(\zeta+\gamma x^Tx')^Q$ $\gamma>0,\zeta\ge0$



Margin为1的10次多项式

SVM + Polynomial Kernel = Polynomial SVM

- Q、γ、ζ 确定了多项式核函数的形式
- 利用核函数可不依赖于$\tilde d$ 获得大间隔分类面

∅ 核函数实现无穷维升维 高斯核$k(x,x')=\exp(-\gamma\|x-x'\|^2)$，$\gamma>0$

为简单起见，考虑一维变量，即：$x=(x)$

$$K(x,x')=\exp(-(x-x')^2)=\exp(-(x)^2)\exp(-(x')^2)\exp(2xx')$$
$$=\exp(-(x)^2)\exp(-(x')^2)\left(\sum_{i=0}^\infty\frac{(2xx')^i}{i!}\right)$$
$$=\sum_{i=0}^\infty\left(\exp(-(x)^2)\exp(-(x')^2)\sqrt{\frac{2^i}{i!}}\sqrt{\frac{2^i}{i!}}(x)^i(x')^i\right)$$
$$\Phi(x)=\exp(-(x)^2)\cdot\left(1,\sqrt{\frac{2^1}{1!}}x,\sqrt{\frac{2^2}{2!}}x^2,\sqrt{\frac{2^3}{3!}}x^3,\dots,\dots\right)$$

工智能与自动化学院

高斯核函数SVM

$$K(x,x')=\exp(-\gamma\|x-x'\|^2) \qquad \gamma>0$$

输入测试样本 x：$g_{SVM}=sign(\sum_{SV}\alpha_ny_nK_\Phi(x_n,x)+b)$
$=sign(\sum_{SV}\alpha_ny_n\exp(-\gamma\|x-x_n\|^2)+b)$

- 以所有支撑向量 x_n 为中心的高斯函数的线性组合
- 也被称之为径向基核函数(Radial Basis Function, RBF)

高斯核函数SVM：通过求取α_n确定所有支撑向量 x_n，构造以支撑向量为中心的高斯函数的线性组合，实现在无穷维空间获得最大间隔分类面