机器学习复习4

2022年7月7日 星期四 13:05

支持向量机 SVM (二)

- kernels 函数

features in large dimensions

eg: $\phi(x)=\begin{bmatrix} x \\ x^2 \\ x^3 \end{bmatrix}$

Def: $K(x,z)=\phi(x)^T\phi(z)$

eg: $K(x,z)=(x^Tz)^2=\left(\sum_{i=1}^n x_iz_i\right)\left(\sum_{j=1}^n x_jz_j\right)$

$=\sum_{i=1}^n\sum_{j=1}^n x_ix_jz_iz_j=\sum_{i=1}^n\sum_{j=1}^n(x_ix_j)(z_iz_j)$

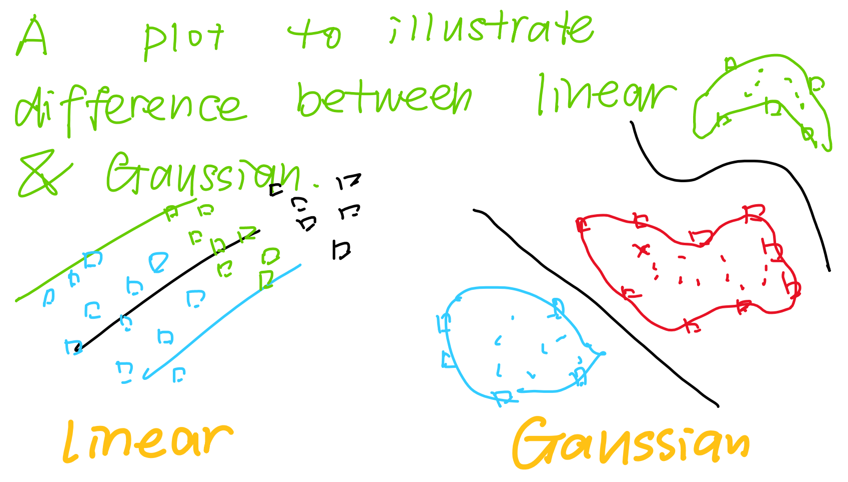
$=\phi(x)^T\phi(z)$ // 可减少时间复杂度，

$O(n^2)$ 仅计算 $(x^Tz)^2$ 就好！

$O(n)$

上述 $\phi(x)=\begin{bmatrix} x_1x_1 \\ x_1x_2 \\ x_1x_3 \\ \vdots \\ x_3x_3 \end{bmatrix}$

A plot to illustrate difference between linear & Gaussian.



对于 $w^Tx+b=\left(\sum_{i=1}^m\alpha_iy^{(i)}x^{(i)}\right)^Tx+b$

$=\sum_{i=1}^m\alpha_iy^{(i)}\langle x^{(i)},x\rangle+b$

只需将 $\langle x^{(i)},x\rangle$ 替换成 $K(x^{(i)},x)$，值判断同上

- 核函数有效性判断

K是有效的核函数 ⇒ 核函数矩阵K是半正定的

? - Mercer 定理

若K是 $\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ 上的映射，如果K为一个有效核函数，那么对训练样例 $\{x^{(1)},x^{(2)},\dots,x^{(m)}\}$ 其相应的核函数矩阵是对称半正定的。

- Regularization & the non-separable case

$\min_{\gamma,w,b}\ \frac{1}{2}\|w\|^2+C\sum_{i=1}^m\xi_i$

s.t. $y^{(i)}(w^Tx^{(i)}+b)\geq 1-\xi_i$

$\xi_i\geq 0$

- Coordinate ascent

eg. $\max_\alpha W(\alpha_1,\alpha_2\dots\alpha_m)$

Loop until convergence:{

For $i=1,\dots,m${

$\alpha_i:=\arg\max_{\hat\alpha_i}W(\alpha_1,\dots\alpha_{i-1},\hat\alpha_i,\alpha_{i+1}\dots\alpha_m)$

}

}

- SMO algorithm (Sequential minimal optimization)

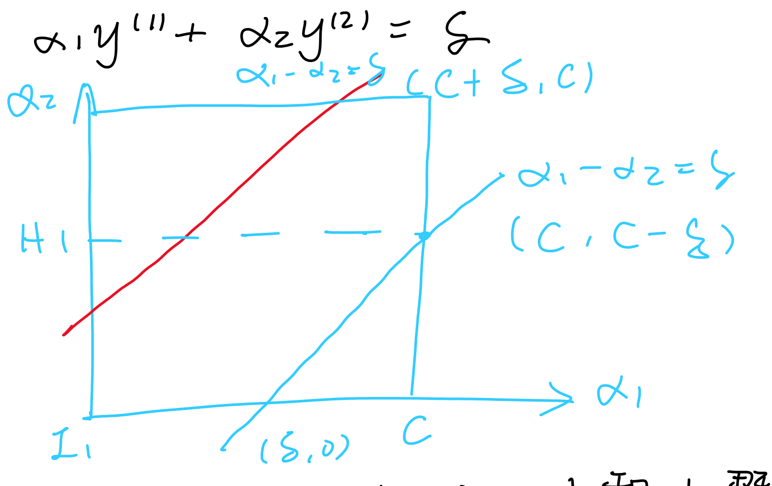
$\max_\alpha\ W(\alpha)=\sum_{i=1}^m\alpha_i-\frac{1}{2}\sum_{i,j=1}^m y^{(i)}y^{(j)}\alpha_i\alpha_j\langle x^{(i)},x^{(j)}\rangle$

s.t. $0\leq\alpha_i\leq C,\ i=1,\cdots,m$

$\sum_{i=1}^m\alpha_iy^{(i)}=0$

$\alpha_1y^{(1)}+\alpha_2y^{(2)}=-\sum_{i=3}^m\alpha_iy^{(i)}$

$\alpha_1y^{(1)}+\alpha_2y^{(2)}=\zeta$



横轴为 α_1，纵轴 α_2，α_1和α_2既要在矩形方框内，也要在直线上，故：

$L=\max(0,\alpha_2-\alpha_1)$，$H=\min(C,C+\alpha_2-\alpha_1)$

同理，当 $y^{(1)}$ 和 $y^{(2)}$ 同号时

$L=\max(0,\alpha_2+\alpha_1-C)$，$H=\min(C,\alpha_2+\alpha_1)$

$$a_2^{new}=\begin{cases} H & \text{if } \alpha_2>H \\ \alpha_2^{new,unclipped} & \text{if } L\leq\alpha_2\leq H \\ L & \text{if } \alpha_2<L \end{cases}$$

$\Rightarrow\alpha_2(K_{11}+K_{22}-2K_{12})=\alpha_2^*(K_{11}+K_{22}-2K_{12})+y_2(u_1-u_2+y_2-y_1)$

$\alpha_2^{new}=\alpha_2+\frac{y_2(E_1-E_2)}{\eta}$

$\alpha_1^{new}=\alpha_1+s(\alpha_2-\alpha_2^{new,clipped})$

⇓

$\vec{w}^{new}=\vec{w}+y_1(\alpha_1^{new}-\alpha_1)\vec{x_1}+y_2(\alpha_2^{new,clipped}-\alpha_2)\vec{x_2}$