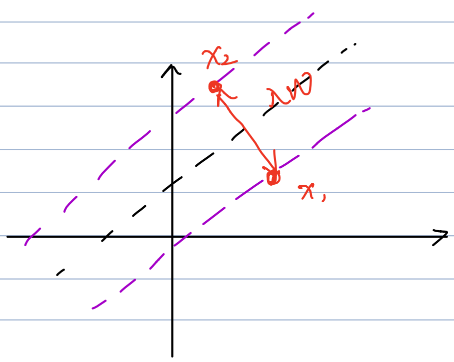
$x - y + 3 = 0$

$(1, -1)\begin{pmatrix} x \\ y \end{pmatrix} + 3 = 0$

$w^T = (1, -1)$，为直线 $x-y+3=0$ 的法向量，对应单位向量为 $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$

$$\begin{cases} w^T x_i + b > 1, & y_i = 1 \\ w^T x_i + b < -1, & y_i = 0 \end{cases}$$

$\Rightarrow y_i(w^T x_i + b) \geq 1$



对 x_2 有 $w^T x_2 + b = 1$

对 x_1 有 $w^T x_1 + b = -1$

对 x_1, x_2 有 $x_2 = x_1 + \lambda w$

$$\begin{cases} w^T(x_1 + \lambda w) + b = 1 \\ w^T x_1 + b = -1 \end{cases}$$

$\therefore \lambda w^T w = 2$，$\lambda = \frac{2}{\|w\|^2}$

λ 即两线间隔，SVM 算法目标最大化间隔。

$$\Rightarrow \begin{cases} \max \|x_1 - x_2\| \\ 1 - y_i(w^T x_i + b) \leq 0 \end{cases}$$

最大化 λ 即最小化 w

$$\Rightarrow \begin{cases} \min \frac{1}{2}\|w\|^2 \\ 1 - y_i(w^T x_i + b) \leq 0 \end{cases}$$

根据拉格朗日定理有（参见拉格朗日 pdf）

$$L = \min_{w,b} \max_{\lambda} \frac{1}{2}\|w\|^2 + \sum_i \lambda_i [1 - y_i(w^T x_i + b)]$$

根据强对偶条件 min max 位置可以互换（参见另一个 pdf）

$$\Leftrightarrow L = \max_{\lambda} \min_{w,b} \frac{1}{2}\|w\|^2 + \sum_i \lambda_i [1 - y_i(w^T x_i + b)]$$

先求 $\min_{w,b}$ 问题.

① $\frac{\partial L}{\partial w} = 0$，$w + \sum_i \lambda_i [-y_i x_i] = 0 \Rightarrow w = \sum_i \lambda_i y_i x_i$

② $\frac{\partial L}{\partial b} = 0$，$\sum_i \lambda_i y_i = 0$

代入原式最终 $L = \max_{\lambda} -\frac{1}{2}\sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_i \lambda_i \Rightarrow \begin{cases} \lambda_i \geq 0 \\ \sum_i \lambda_i y_i = 0 \end{cases}$

具体推算： $L = \max\limits_{\lambda} \boxed{\frac{1}{2} w^T w} + \sum_i \lambda_i \underbrace{- \sum_i \lambda_i y_i b}_{0} \boxed{- \sum_i \lambda_i y_i w^T x_i}$

$= \max\limits_{\lambda} -\frac{1}{2} w^T w + \sum_i \lambda_i$ $\qquad w = \sum_i \underbrace{\lambda_i y_i}_{标量} \underbrace{x_i}_{向量}$

$= \max\limits_{\lambda} -\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_i \lambda_i$

$\begin{cases} \lambda_i \geq 0 \\ \sum_i \lambda_i y_i = 0 \end{cases}$

这里已经求完了 $\min\limits_{w,b}$ 问题，下一步求 $\max\limits_{\alpha}$

根据 SMO 算法 $\Rightarrow \lambda^*$

对应的 $w = \sum_i \lambda_i y_i x_i$ ， $y_i [1 - (w^T x_i + b)] = 0$ 求出 w, b 结束

简单理解支持向量：

$L = \frac{1}{2} \|w\|^2 + \sum_i \lambda_i \underline{[1 - y_i (w^T x_i + b)]}$

$\begin{cases} \text{在线上，即边界上} \Rightarrow y_i(w^T x_i + b) = 1 \\ \text{在线外，即相当于无约束，} \lambda_i = 0 \end{cases}$

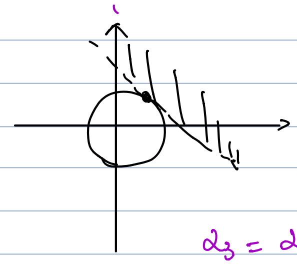
这里可知求出 $\lambda_i \neq 0$ 的点对应的向量就是支持向量。

手推实例： $x_1 = (3, 3)$ $y_1 = 1$

$x_2 = (4, 3)$ $y_2 = 1$

$x_3 = (1, 1)$ $y_3 = -1$

$\begin{cases} \min (w_1^2 + w_2^2) \\ 3w_1 + 3w_2 + b \geq 1 \\ 4w_1 + 3w_2 + b \geq 1 \\ -w_1 - w_2 - b \geq 1 \end{cases}$ $\qquad 2w_1 + 2w_2 \geq 2$



图示 $w_1 = w_2 = \frac{1}{2}$ 为极值。

$\alpha_3 = \alpha_1 + \alpha_2$

$y_i y_j$	1	2	3
1	1	1	-1
2	1	1	-1
3	-1	-1	1

$x_i x_j$	1	2	3
1	18	21	6
2	21	25	7
3	6	7	2

根据 svm ， $L = \max\limits_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$ ， $\sum_i \alpha_i y_i = 0$

$L = \max\limits_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} [\alpha_1^2 \cdot 18 + \alpha_1 \alpha_2 \cdot 21 \cdot 2 + \alpha_1 \cdot \alpha_3 \cdot (-6) \cdot 2 + \alpha_2^2 \cdot 25 + \alpha_2 \cdot \alpha_3 \cdot (-7) \cdot 2 + \alpha_3^2 \cdot 2]$

$= \max\limits_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 - [9\alpha_1^2 + 21\alpha_1\alpha_2 - 6\alpha_1\alpha_3 + \frac{25}{2}\alpha_2^2 - 7\alpha_2\alpha_3 + \alpha_3^2]$

$L = \max\limits_{\alpha} 2\alpha_1 + 2\alpha_2 - 9\alpha_1^2 - 21\alpha_1\alpha_2 + 6\alpha_1^2 + 6\alpha_1\alpha_2 - \frac{25}{2}\alpha_2^2 + 7\alpha_2^2 + 7\alpha_1\alpha_2 - \alpha_1^2 - \alpha_2^2 - 2\alpha_1\alpha_2$

$= \max\limits_{\alpha} -4\alpha_1^2 - \frac{13}{2}\alpha_2^2 - 10\alpha_1\alpha_2 + 2\alpha_1 + 2\alpha_2$

$$\frac{\partial L}{\partial \alpha_1} = -8\alpha_1 - 10\alpha_2 + 2 = 0$$

$$\frac{\partial L}{\partial \alpha_2} = -13\alpha_2 - 10\alpha_1 + 2 = 0$$

$\alpha_1 = 1.5,\ \alpha_2 = -1$ × $\longrightarrow$ $\begin{cases} (0,\alpha_2) & -\frac{13}{2}\alpha_2^2 + 2\alpha_2 \quad \alpha_2 = \frac{2}{13} \\ (\alpha_1, 0) & \alpha_1 = \frac{1}{4}. \end{cases}$

$L_{\alpha_1=\frac14} = -4\cdot\frac{1}{16} + \frac12 = \frac14$ 上面大，选 $\frac14$

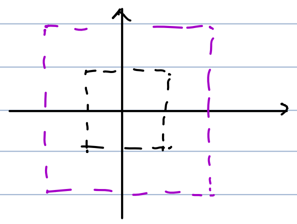
$L_{\alpha_2=\frac{2}{13}} = -\frac{2}{13} + \frac{4}{13} = \frac{2}{13}$

$\alpha_1 = \frac14,\quad \alpha_2 = 0,\quad \alpha_3 = \frac14$

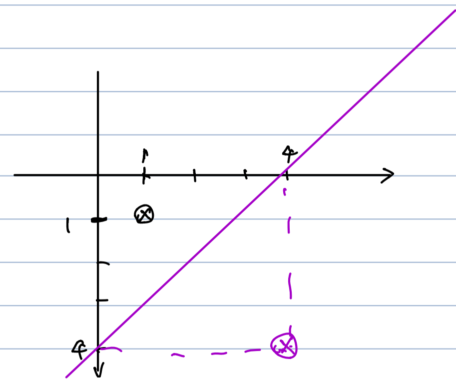
$$w^* = \sum_i \alpha_i y_i x_i = \frac14\begin{pmatrix}3\\3\end{pmatrix} - \frac14\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}\frac12\\ \frac12\end{pmatrix}$$

$$w^* x_1 + b = 1 \Rightarrow \begin{pmatrix}\frac12\\ \frac12\end{pmatrix}^T\begin{pmatrix}3\\3\end{pmatrix} + b = 1,\quad b = -2.$$

核方法



$\Rightarrow$



$R^2 \Longrightarrow R^3$

$(x_1, x_2) \Longrightarrow (z_1, z_2, z_3)$

$z_1 = x_1^2,\quad z_2 = \sqrt{2}\,x_1 x_2,\quad z_3 = x_2^2$

$(1, 1)$	$1,\ \sqrt2,\ 1$
$(-1, 1)$	$1,\ -\sqrt2,\ 1$
$(-1, -1)$	$1,\ \sqrt2,\ 1$
$(1, -1)$	$1,\ -\sqrt2,\ 1$
$(2, 2)$	$4,\ 4\sqrt2,\ 4$
$(-2, 2)$	$4,\ -4\sqrt2,\ 4$
$(-2, -2)$	$4,\ 4\sqrt2,\ 4$
$(2, -2)$	$4,\ -4\sqrt2,\ 4$

核方法能解决一部分线性不可分问题

增加特征维度

$$L = \max_{\alpha} \sum_i \alpha - \frac12 \sum_i\sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\begin{cases} \sum_i \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases}$$

引入核方法 $L = \max_{\alpha} \sum_i \alpha - \frac12 \sum_i\sum_j \alpha_i \alpha_j y_i y_j \phi^T(x_i)\phi(x_j)$

高维求内积消耗大，找一个核函数使得 $K(x_i, x_j) = \phi^T(x_i)\phi(x_j)$

核函数对应核矩阵应是半正定，符合 mercer 定理.

线性核： $x_i^T x_j$

多项式： $(x_i^T x_j)^d$

高斯核： $\exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2}\right)$

拉普拉斯核： $\exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right)$

sigmoid 核： $\tanh(\beta x_i^T x_j + \theta)$ $\quad \frac{e^x - e^{-x}}{e^x + e^{-x}}$

软间隔允许分错

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_{0/1}\left(1 - y_i(w^T x_i + b)\right)$$

C 无穷大 $\Rightarrow$ 不允许出错.

$$l_{0/1}(z) = \begin{cases} 1 & z > 0 \\ 0 & z \le 0 \end{cases}$$

$l_{0/1}$ 可被换成：

hinge loss : $\max(0, 1-z)$

指数 loss : $\exp(-z)$

对数 loss : $\log[1+\exp(-z)]$

改写一下

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\varepsilon_i$$

$$\begin{cases} y_i(w^T x_i + b) \ge 1 - \varepsilon_i \\ \varepsilon_i \ge 0 \qquad -\varepsilon_i \le 0 \end{cases}$$

$$\Rightarrow \mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\varepsilon_i + \sum_{i=1}^{m}\alpha_i[1-\varepsilon_i - y_i(w^T x_i + b)] - \sum_{i=1}^{m}\mu_i\varepsilon_i$$

$$\begin{cases} \alpha_i \ge 0, \quad y_i f(x_i) \ge 1-\varepsilon_i, \quad \alpha_i[1-\varepsilon_i - y_i f(x_i)] = 0 \\ \mu_i \ge 0, \quad \varepsilon_i \ge 0, \quad \mu_i\varepsilon_i = 0 \end{cases}$$

$$\Rightarrow \begin{cases} w = \sum_{i=1}^{m}\alpha_i y_i x_i \\ \sum \alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \\ C_i = \alpha_i + \mu_i \end{cases}$$

$\alpha_i = 0$ 时 不影响

$\alpha_i \ne 0$ 时 $y_i f(x_i) = 1 - \varepsilon_i$

$\alpha_i = C$ 时 $\mu_i = 0$ $\begin{cases} \varepsilon_i \le 1, \text{ 在分隔线里面} \\ \varepsilon_i > 1, \text{ 分错了, 此时 } 1-\varepsilon_i < 0 \end{cases}$

$\alpha_i < C$ 时 $\mu_i > 0$, $\varepsilon_i = 0$

KLDA: $h(x) = \sum_{i=1}^{m}\alpha_i k(x, x_i)$.

$$\hat{\mu}_0 = \frac{1}{m_0}K_{10}, \quad \hat{\mu}_1 = \frac{1}{m_1}K_{11}, \quad M = (\hat{\mu}_0 - \hat{\mu}_1)(\hat{\mu}_0 - \hat{\mu}_1)^T, \quad N = KK^T - \sum_{i=0}^{1} m_i\hat{\mu}_i\hat{\mu}_i^T$$

$$\max_{\alpha} J(\alpha) = \frac{\alpha^T M\alpha}{\alpha^T N\alpha}$$

SVR: $\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_\varepsilon(f(x_i) - y_i)$ $\Rightarrow$ $l_\varepsilon(z) = \begin{cases} 0, & |z| \le \varepsilon \\ |z| - \varepsilon, & \text{otherwise} \end{cases}$

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\varepsilon_i + \hat{\varepsilon}_i)$$

$$\begin{cases} f(x_i) - y_i \le \varepsilon + \varepsilon_i \\ y_i - f(x_i) \le \varepsilon + \hat{\varepsilon}_i \\ \varepsilon_i \ge 0, \ \hat{\varepsilon}_i \ge 0, \quad i = 1, \dots, m \end{cases}$$

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) - \sum_{i=1}^{m}\mu_i\xi_i - \sum_{i=1}^{m}\hat{\mu}_i\hat{\xi}_i$$

$$+ \sum_{i=1}^{m} a_i\left[f(x_i) - y_i - \varepsilon - \xi_i\right] + \sum_{i=1}^{m}\hat{a}_i\left[y_i - f(x_i) - \varepsilon - \hat{\xi}_i\right]$$

$$\begin{cases} \mu_i \geq 0, \quad \hat{\mu}_i \geq 0, \quad \mu_i\xi_i = 0, \quad \hat{\mu}_i\hat{\xi}_i = 0, \quad \xi_i \geq 0, \quad \hat{\xi}_i \geq 0 \\ a_i \geq 0, \quad \hat{a}_i \geq 0, \quad f(x_i) - y_i - \varepsilon - \xi_i \leq 0, \quad y_i - f(x_i) - \varepsilon - \hat{\xi}_i \leq 0 \\ a_i\left[f(x_i) - y_i - \varepsilon - \xi_i\right] = 0, \quad \hat{a}_i\left[y_i - f(x_i) - \varepsilon - \hat{\xi}_i\right] \end{cases}$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{m}(\hat{a}_i - a_i)x_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow 0 = \sum_{i=1}^{m}(\hat{a}_i - a_i)$$

$$C = \mu_i + a_i$$

$$C = \hat{\mu}_i + \hat{a}_i$$

代入 $L = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{a}_i - a_i)(\hat{a}_j - a_j)x_i^T x_j + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\mu_i\xi_i + C\sum_{i=1}^{m}\hat{\xi}_i - \sum_{i=1}^{m}\hat{\mu}_i\hat{\xi}_i$

$$+ \sum_{i=1}^{m} a_i w_i^T x_i + \sum_{i=1}^{m}\hat{a}_i(-w_i^T x_i) + \sum_{i=1}^{m} -a_i\xi_i + \sum_{i=1}^{m} -\hat{a}_i\hat{\xi}_i$$

$$+ \sum_{i=1}^{m} a_i(-y_i - \varepsilon + b) + \sum_{i=1}^{m}\hat{a}_i(y_i - \varepsilon - b)$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{a}_i - a_i)(\hat{a}_j - a_j)x_i^T x_j - \sum_{i=1}^{m}(\hat{a}_i - a_i)w_i^T x_i$$

$$+ b\sum_{i=1}^{m}(a_i - \hat{a}_i) + \sum_{i=1}^{m} y_i(\hat{a}_i - a_i) - \varepsilon(\hat{a}_i + a_i)$$

$$= \sum_{i=1}^{m} y_i(\hat{a}_i - a_i) - \varepsilon(\hat{a}_i + a_i) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{a}_i - a_i)(\hat{a}_j - a_j)x_i^T x_j$$

$$\begin{cases} \cancel{C = \mu_i + a_i} \\ \cancel{C = \hat{\mu}_i + \hat{a}_i} \\ \sum_{i=1}^{m}(\hat{a}_i - a_i) = 0 \\ 0 \leq a_i, \hat{a}_i \leq C \end{cases} \Rightarrow \begin{cases} a_i\left[f(x_i) - y_i - \varepsilon - \xi_i\right] = 0, \quad \hat{a}_i\left[y_i - f(x_i) - \varepsilon - \hat{\xi}_i\right] = 0 \\ \cancel{a_i\hat{a}_j = 0, \quad \mu_i\hat{\mu}_j = 0} \\ \cancel{f(x_i) - y_i \leq \varepsilon + \xi_i, \quad y_i - f(x_i) \leq \varepsilon + \hat{\xi}_i} \\ \cancel{\mu_i \geq 0, \quad \mu_j \geq 0, \quad a_i \geq 0, \quad a_j \geq 0} \\ a_i\hat{a}_j = 0, \quad \xi_i\hat{\xi}_j = 0, \quad (C - a_i)\xi_i = 0, \\ (C - \hat{a}_i)\hat{\xi}_i = 0. \end{cases}$$

$$w = \sum_{i=1}^{m}(\hat{a}_i - a_i)x_i$$

$$f = \sum_{i=1}^{m}(\hat{a}_i - a_i)x_i^T x + b$$

$$\underline{b = f(x) + \varepsilon - \sum_{i=1}^{m}(\hat{a}_i - a_i)x_i^T x}$$

当 $0 < a < C$, $\xi_i = 0$

改写核方法

$$w = \sum_{i=1}^{m}(\hat{a}_i - a_i)\phi(x_i)$$

$$f = \sum_{i=1}^{m}(\hat{a}_i - a_i)k(x, x_i) + b.$$