SVM

- 线性可分支持向量机　线性可分　硬间隔最大化
- 线性支持向量机　近似线性可分　软间隔最大化
  - ↳有异常点, outliers
- 非线性支持向量机　线性不可分　利用核函数在特征空间隐式学习线性支持向量机

线性可分支持向量机

分离超平面　$w^* x + b^* = 0$

决策函数　$f(x) = \mathrm{sign}(w^* x + b^*)$

一个点到超平面距离的度量

- 函数间隔　$y^i (w x^i + b)$
- 几何间隔　$\dfrac{y^i (w x^i + b)}{\|w\|}$

硬间隔最大化 → 几何间隔

定义训练集到超平面的几何间隔

$$\gamma = \min_{i=1,2,\cdots N} \frac{y^i (w x^i + b)}{\|w\|}$$ (所有点中几何间隔最小的)

函数间隔

$$\hat{\gamma} = \min_{i=1,2,\cdots N} y^i (w x^i + b)$$

定义问题

① $\begin{cases} \max\limits_{w,b} \gamma \\ s.t \ \boxed{\dfrac{y_i(Wx_i+b)}{\|w\|} \geq \gamma} \quad i=1,2,\cdots N \end{cases}$

$\Updownarrow \hat{\gamma}$

② $\begin{cases} \max\limits_{w,b} \dfrac{\hat{\gamma}}{\|w\|} \\ s.t \ y_i(Wx_i+b) \geq \hat{\gamma} \quad i=1,2,\cdots N \end{cases}$

③ 取$\hat{\gamma}=1$ $\max\limits_{w,b} \dfrac{1}{\|w\|} \Leftrightarrow \min\limits_{w,b} \frac{1}{2}\|w\|^2$

$\begin{cases} \boxed{\min\limits_{w,b} \frac{1}{2}\|w\|^2} \\ s.t \ y_i(Wx_i+b) \geq 1 \end{cases}$ ⇒表示的是间隔最大化

有了问题的形式，接下来是如何求解该问题了
典型的有约束最优化问题，转化为对偶问题

学习算法

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i (y_i(Wx_i+b)-1)$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i y_i (Wx_i+b) + \sum_{i=1}^{N} \alpha_i$$

$(\alpha_i \geq 0)$

$$\max_{\alpha} \min_{w,b} L(w,b,\alpha)$$

求解 $\min\limits_{w,b} L(w,b,\alpha)$

$$\frac{\partial L(w,b,\alpha)}{\partial w} = W - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0$$

通过求解 $\min\limits_{w,b} L(w,b,\alpha)$，得到了

$$w^* = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\min_{w,b} L(w,b,\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i \cdot x_j - \sum_{i=1}^{N} \alpha_i y_i \left[\sum_{j=1}^{N} \alpha_j y_j (x_j \cdot x_i) + b\right] + \sum_{i=1}^{N} \alpha_i$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i \cdot x_j - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \boxed{- b\sum_{i=1}^{N} \alpha_i y_i} + \sum_{i=1}^{N} \alpha_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i.$$

求解 $\max\limits_{\alpha} (\min\limits_{w,b} L(w,b,\alpha))$

$$\max_{\alpha} (\min_{w,b} L(w,b,\alpha)) =$$

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i$$

$$\text{s.t.} \ \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

通过最优化算法求解上述问题，可得最优解 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$
但是我们只知道 w^* 的形式，代入 α^* 就有了 w^*，不知道 b^* 的形式，怎么办？
利用KKT条件求解.

回到原问题

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{N}\alpha_i(y_i(wx_i+b)-1)$$

$$=\frac{1}{2}\|w\|^2-\sum_{i=1}^{N}\alpha_i y_i(wx_i+b)+\sum_{i=1}^{N}\alpha_i$$

$$(\alpha_i\geq 0)$$

可以列出KKT条件

$$\begin{cases}\nabla_w L(w,b,\alpha)=w^*-\sum_{i=1}^{N}\alpha_i^* y_i x_i=0\\ \nabla_b L(w,b,\alpha)=-\sum_{i=1}^{N}\alpha_i^* y_i=0\\ \alpha_i^*[y_i(wx_i+b)-1]=0\\ y_i(wx_i+b)\geq 1\\ \alpha_i\geq 0\end{cases}$$

必有$\alpha_j>0$（若α_i全为0，则$w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i=0$），$\alpha_j>0$，则$y_j(wx_j+b)-1=0$

$$b=\frac{1}{y_j}-w\cdot x_j \qquad y_i\in\{1,-1\} \Rightarrow \frac{1}{y_j}=y_j$$

$$b=y_j-w\cdot x_j=y_j-\sum_{i=1}^{N}\alpha_i^* y_i(x_i\cdot x_j).$$

至此，我们得到了线性可分支持向量机的形式 具体求解算法参照 P123.

注意：我们可以看到，$\alpha_i\geq 0$，只有当$y_i(wx_i+b)=1$时，才有$\alpha_i\neq 0$.

而w,b也就仅仅取决于这些$\alpha_i\neq 0$的点.

这些$\alpha_i\neq 0$的点也就是支持向量，w和b仅取决于这些支持向量.

线性支持向量机

如果数据集近似于线性可分，也就是有些异常点，那么需要用软间隔最大化.

软间隔最大化

$$① \begin{cases}\max_{w,b,\xi}\ \gamma+C\sum_{i=1}^{N}\xi_i\\ \text{s.t. } \frac{y_i(w\cdot x_i+b)}{\|w\|}\geq\gamma-\xi_i\end{cases}$$

→ 也就是对于部分点，我们允许不满足大于 γ 间隔 的条件

②
$$\begin{cases} \max\limits_{w,b,\xi} \ r + C\sum_{i=1}^{N}\xi_i \\ s.t \quad y_i(w\cdot x_i+b) \geq \|w\| r - \|w\|\xi_i \end{cases}$$

$$\begin{cases} \max\limits_{w,b,\xi} \ \frac{\hat{r}}{\|w\|} + C\sum_{i=1}^{N}\xi_i \\ s.t. \quad y_i(w\cdot x_i+b) \geq \hat{r} - \|w\|\xi_i \end{cases}$$

取 $\hat{r}=1$

$$\begin{cases} \max\limits_{w,b,\xi} \ \frac{1}{\|w\|} + C\sum_{i=1}^{N}\xi_i \\ s.t \quad y_i(w\cdot x_i+b) \geq 1 - \|w\|\xi_i . \end{cases}$$

$$\begin{cases} \min\limits_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ s.t \quad y_i(w\cdot x_i+b) \geq 1 - \xi_i . \\ \quad\quad \xi_i \geq 0 . \end{cases}$$

分离超平面 $w\cdot x+b=0$

决策函数 $f(x)=\mathrm{sign}(w\cdot x+b)$

学习算法

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i[1-\xi_i-y_i(w\cdot x_i+b)] - \sum_{i=1}^{N}\mu_i\xi_i .$$

$$\max_{\alpha,\mu}\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$$

求解 $\min\limits_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$

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

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N}\alpha_i y_i = 0 \qquad \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0$$

因此我们知道

$$\begin{cases} w^* = \sum_{i=1}^{N} \alpha_i y_i x_i \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \\ C - \alpha_i - \mu_i = 0 \end{cases}$$

$$\min_{w,b,\xi} L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + C\sum_{i=1}^{N}\xi_i$$

$$+ \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i \xi_i - \sum_{i=1}^{N}\mu_i \xi_i \rightarrow 0$$

$$- \sum_{i=1}^{N}\alpha_i y_i \left(\sum_{j=1}^{N}\alpha_j y_j (x_j \cdot x_i) + b\right)$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$- b\sum_{i=1}^{N}\alpha_i y_i + \sum_{i=1}^{N}\alpha_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

与线性可分SVM的形式相同.

$$\max_{\alpha,\mu}\min_{w,b,\xi} L = \max_{\alpha} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{N}\alpha_i y_i = 0 \\ C - \alpha_i - \mu_i = 0 \\ \alpha_i \geq 0 \\ \mu_i \geq 0 \end{cases} \Rightarrow 0 \leq \alpha_i \leq C$$

同样,利用最优化算法求解上述问题,可以得到一个α^*.
我们仍然只是知道w^*的形式,缺一个b^*的形式.

KKT条件

$$L(w,b,\xi,\alpha,\mu)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i+\sum_{i=1}^{N}\alpha_i[1-\xi_i-y_i(w\cdot x_i+b)]-\sum_{i=1}^{N}\mu_i\xi_i$$

$$0\le\alpha_i\le C$$

$$\boxed{\begin{aligned}&\nabla_w L=w-\sum_{i=1}^{N}\alpha_i^* y_i x_i=0\\&\nabla_\xi L=C-\alpha^*-\mu^*=0\\&\nabla_b L=\sum_{i=1}^{N}\alpha_i y_i=0\\&\alpha_i^*[y_i(w^*\cdot x_i+b^*)-1+\xi_i^*]=0\\&\mu_i^*\xi_i^*=0\\&\xi_i^*\geq 0\\&w^* x_i+b^*-1+\xi_i\geq 0\end{aligned}}$$

存在 $0<\alpha_i^*<C$，则 $y_i(w^* x_i+b)-1+\xi_i^*=0$

$\because C-\alpha_i^*-\mu_i^*=0$

$\therefore \mu_i^*\neq 0$

$\therefore \xi_i^*\neq 0$

则 $y_i(w^* x_i+b)-1=0$

$$b^*=y_i-\sum_{j=1}^{N}\alpha_j^* y_j(x_j\cdot x_i)$$

具体求解算法见P129.

$\alpha_i^*=0$ 则 $y_i(w^* x_i+b)\geq 1$ ⇒ 正确分类 非支持向量

$0<\alpha_i^*<C$ 则 $y_i(w^* x_i+b)=1$ ⇒ 正确分类 间隔边界上的支持向量

$\alpha_i^*=C$, ⇒ $0<\xi_i^*<1$ $y_i(w^*\cdot x_i+b)<1$ 正确分类，间隔边界与超平面之间的支持向量

$\mu_i^*=0$

$$\begin{cases} \alpha_i^* = 1 & y_i(w^* \cdot x_i + b) = 0 \quad 超平面上 \\ \alpha_i^* > 1 & y_i(w^* \cdot x_i + b) < 0 \quad 分类错误 \end{cases}$$

非线性支持向量机

决策函数：$f(x) = \mathrm{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x, x_i) + b^*\right)$

这部分书上讲的很好，看书吧！

SMO

到此为止，对于支持向量机，我们只剩下一个问题没有解决了！

那就是如何最优化这个问题？

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i$$

线性支持向量机，写成一个一般的形式

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

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \le \alpha_i \le C$$

适用线性和非线性支持向量机

完整 SMO 算法：（提前注意的有N个x_i，就有N个α_i，x_i与α_i是一一对应的）

输入：训练数据集 $T = \{(x_1, y_1), \cdots (x_N, y_N)\}$，$y \in \{-1, +1\}$，精度 ε

输出：近似解 $\hat{\alpha}$

(1) 初始化 α 为 0，

(2) 对所有的样本点计算 E_i：

$$E_i = \left(\sum_{j=1}^{N} \alpha_j y_j K(x_j \cdot x_i) + b\right) - y_i$$

(2) 选择α_1, α_2来独地优化这两个变量

(1) 首先循环遍历所有满足条件$0<\alpha_i<C$的样本点，检验它们是否满足KKT条件，取一个不满足KKT条件的α_i作为α_1.

(2) 若所有$0<\alpha_i<C$的样本点都满足KKT条件，那么遍历所有数据点，取一个不满足KKT条件的α_i作为α_1

满足KKT条件

$$\begin{cases} \alpha_i=0 \Leftrightarrow y_i g(x_i) \geq 1 \\ 0<\alpha_i<C \Leftrightarrow y_i g(x_i)=1 \\ \alpha_i=C \Leftrightarrow y_i g(x_i)<1 \end{cases}$$

$$g(x_i)=\sum_{j=1}^{N} \alpha_j y_j K(x_j, x_i)+b$$

(3) 若$E_1>0$，则选择最小的E_i作为E_2，相应的α_i作为α_2.
若$E_1<0$，则选择最大的E_i作为E_2，相应的α_i作为α_2.

(4) 计算更新后的α_1, α_2即α_1^{new} α_2^{new}

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

$$\eta=K_{11}+K_{22}-2K_{12}.$$

$$L=\max(0, \alpha_2^{old}-\alpha_1^{old}) \quad H=\min(C, C+\alpha_2^{old}-\alpha_1^{old})$$

$$\alpha_2^{new}=\begin{cases} H & \alpha_2^{new,unc}>H \\ \alpha_2^{new,unc} & L\leq\alpha_2^{new,unc}\leq H \\ L & \alpha_2^{new,unc}<L \end{cases}$$

$$\alpha_1^{new}=\alpha_1^{old}+y_1y_2(\alpha_2^{old}-\alpha_2^{new})$$

4)计算 b^{new}

$$b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

$$b_2^{new} = -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

若 $0<\alpha_1^{new}<C$ 且 $0<\alpha_2^{new}<C$，$b^{new} = b_1^{new}$ 或 $b^{new} = b_2^{new}$

其余情况 $b^{new} = \frac{1}{2}(b_1^{new} + b_2^{new})$