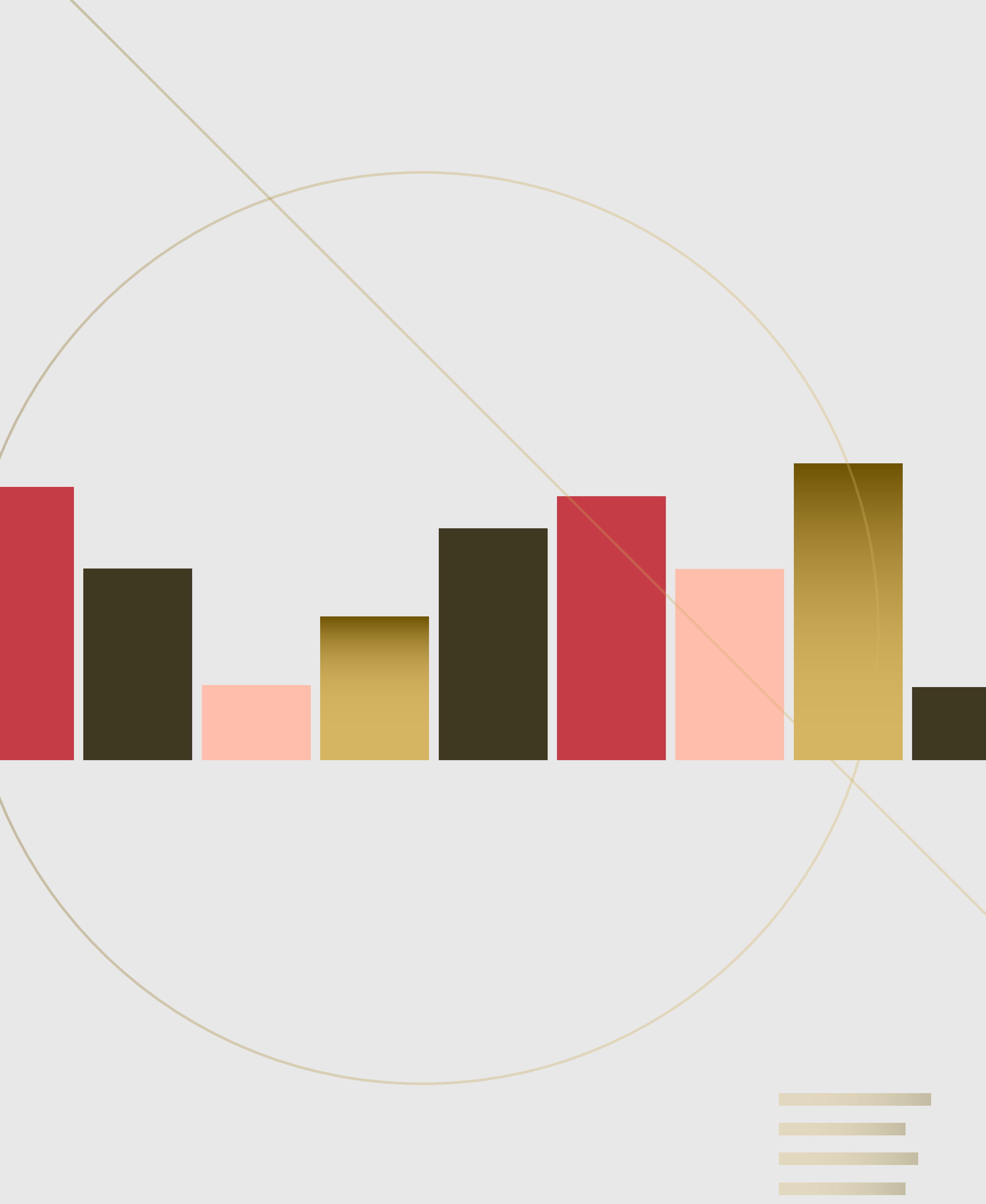

给定训练样本集 $D=\{(x_1,y_1),(x_2,y_2),\cdots(x_n,y_n)\}$，$y_i\in\{-1,+1\}$，分类学习的基本思想就是基于训练样本D在样本空间中找到一个划分超平面，将不同类别的样本分开。

在样本空间中，划分超平面可表示为

$$w^T x + b = 0$$

其中 $w=(w_1;w_2;\cdots;w_d)$ 为法向量，决定超平面的方向；b为位移项，决定超平面与原点之间的距离。将超平面记为 (w,b)，那么样本空间中任意一点到超平面的距离为

$$r = \frac{|w^T x+b|}{\|w\|}$$

假设此时有一个超平面 (w',b') 能使样本正确分类，则对于 $(x_i;y_i)\in D$ 有：

$$\begin{cases} (w')^T x_i + b' > 0 & y_i = +1 \\ (w')^T x_i + b' < 0 & y_i = -1 \end{cases} \quad ①$$

根据样本点与超平面之间的几何间隔关系，则上式可改为

$$\begin{cases} (w')^T x_i + b' \geq +\varepsilon, & y_i = +1 \\ (w')^T x_i + b' \leq -\varepsilon, & y_i = -1 \end{cases}$$ ε为大于0的常数 ②

$$\Rightarrow \begin{cases} \left(\frac{w'}{\varepsilon}\right)^T x_i + \left(\frac{b'}{\varepsilon}\right) \geq +1, & y_i = +1 \\ \left(\frac{w'}{\varepsilon}\right)^T x_i + \left(\frac{b'}{\varepsilon}\right) \leq -1, & y_i = -1 \end{cases}$$

令 $w=\frac{w'}{\varepsilon}$，$b=\frac{b'}{\varepsilon}$，则 $\begin{cases} wx_i + b \geq +1, & y_i = +1 \\ wx_i + b \leq -1, & y_i = -1 \end{cases}$ ③

在训练样本中，距离超平面最近的几个样本能使③式等号成立，则这些样本被称为"支持向量"，那么两个异类支持向量到超平面的距离为

$$\gamma = \frac{2}{\|w\|}$$

此时 α 被称为间隔

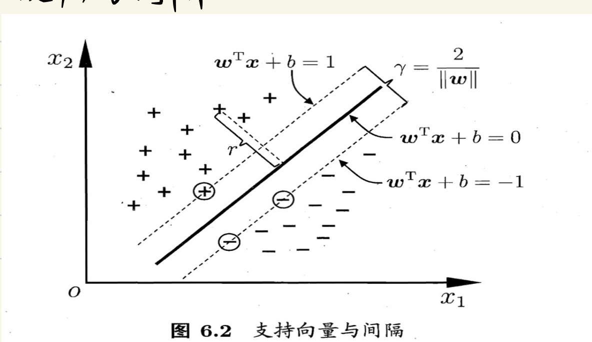

图 6.2 支持向量与间隔

对于分类，我们的目标就是找到最大间隔的划分超平面，此时转化为求解 w，b，使得 γ 最大，约束模型为

$$\max_{w,b}\frac{2}{\|w\|} \quad s.t.\ \ y_i(w^Tx_i+b)\geq 1 \quad i=1,2,\cdots,n. \qquad ④$$

要使 γ 最大，则等价于 $\|w\|^2$ 最小，于是④可改写为

$$\min_{w,b}\|w\|^2 \quad s.t.\ \ y_i(w^Tx_i+b)\geq 1,\ \ i=1,2,\cdots,n \qquad ⑤$$

为求解⑤式可使用拉格朗日乘数法可得到其对偶问题，故有

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i(1-y_i(w^Tx_i+b)) \qquad ⑥$$

令 $\frac{\partial L}{\partial w}=0$ $\frac{\partial L}{\partial b}=0$，有

$$\begin{aligned}L(w,b,\alpha)&=\frac{1}{2}w^Tw+\sum_{i=1}^{m}(\alpha_i-\alpha_iy_iw^Tx_i-\alpha_iy_ib)\\&=\frac{1}{2}w^Tw+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_iy_iw^Tx_i-\sum_{i=1}^{m}\alpha_iy_ib \qquad ⑦\end{aligned}$$

$$\begin{cases}\frac{\partial L(w,b,\alpha)}{\partial w}=w-\sum_{i=1}^{m}\alpha_iy_ix_i=0\\ \frac{\partial L(w,b,\alpha)}{\partial b}=-\sum_{i=1}^{m}\alpha_iy_i=0\end{cases}$$

$$\Rightarrow\begin{cases}w=\sum_{i=1}^{m}\alpha_iy_ix_i\\ 0=\sum_{i=1}^{m}\alpha_iy_i\end{cases} \qquad ⑧$$

将⑧代入⑦中，有

$$\begin{aligned}\min_{w,b}L(w,b,\alpha)&=\frac{1}{2}w^Tw+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_iy_iw^Tx_i-\sum_{i=1}^{m}\alpha_iy_ib\\&=\frac{1}{2}w^T\cdot\sum_{i=1}^{m}\alpha_iy_ix_i+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}\alpha_iy_iw^Tx_i\\&=-\frac{1}{2}w^T\sum_{i=1}^{m}\alpha_iy_ix_i+\sum_{i=1}^{m}\alpha_i\\&=-\frac{1}{2}\left(\sum_{i=1}^{m}\alpha_iy_ix_i\right)^T\left(\sum_{i=1}^{m}\alpha_iy_ix_i\right)+\sum_{i=1}^{m}\alpha_i\\&=-\frac{1}{2}\sum_{i=1}^{m}\alpha_iy_ix_i^T\sum_{i=1}^{m}\alpha_iy_ix_i+\sum_{i=1}^{m}\alpha_i\\&=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_jx_i^Tx_j\end{aligned}$$

则⑤的对偶问题为

$$\max_{\alpha}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_jx_i^Tx_j$$

$$s.t.\ \sum_{i=1}^{m}\alpha_iy_i=0 \qquad \alpha_i\geq 0$$

在求解出 α 后，则超平面方程为

$$f(x)=\sum_{i=1}^{m}\alpha_iy_ix_i^Tx+b$$