第2章 统计模式识别中的几何方法

2.1 统计分类的基本思想

简单模式指：

- 模式具有表征其类别的类属性特征。
- 属于同一个模式类的各个模式，其类属性特征在特征空间中应组成某种程度上的一个集群区域。
- 不同模式类的类属性特征在特征空间中组成的区域应该是彼此分离的。

2.2 模式的相似性度量和最小距离分类器

常见度量函数

- Minkowsky距离： $d(X,Y)=\left[\sum\limits_{i=1}^{n}|x_i-y_i|^\lambda\right]^{\frac{1}{\lambda}}$
- Manhattan距离， $d(X,Y)=\sum\limits_{i=1}^{n}|x_i-y_i|$ ，就是第一个在 $\lambda=1$ 时的特例。
- Cityblock距离， $d(X,Y)=\sum\limits_{i=1}^{n}\omega_i|x_i-y_i|$ ，对上一个加权
- Euclidean距离， $d(X,Y)=\left[\sum\limits_{i=1}^{n}|x_i-y_i|^2\right]^{\frac{1}{2}}$ ，称为欧氏距离。
- Camberra距离， $d(X,Y)=\sum\limits_{i=1}^{n}\frac{|x_i-y_i|}{|x_i+y_i|}$
- Mahalanobis距离， $d^2(X,M)=(X-M)^T\Sigma^{-1}(X-M)$ ，其中M为某一个模式类别的均值向量， Σ 为相应模式类别的协方差矩阵。

在扇形分布中，也常见： $S(X,Y)=\cos\theta=\frac{X^TY}{||X||\,||Y||}$

基于标准样本的距离分类器

假设每个模式类都有标准样本 $M_1,M_2,\cdots,M_n$ 表示，那么若 $d(X,M_i)=\min\limits_{j}\{d(X,M_j)\}\Rightarrow X\in\omega_i$

若采用欧氏距离，记 $g_j(X)=M_j^TX-\frac{1}{2}M_j^TM_j$ ，则判决条件改为 $g_i(X)>g_j(X)$ ；也可以定义增广向量（P24）

基于分散样本的最小距离分类器

- **平均样本法** 定义标准样本为均值。简单，但是没有考虑到样本散布对分类器性能的影响，分类性能一般不理想。
- **平均距离法** 定义距离为到各点距离的平均。有一定抗噪声能力，但是计算量大，需要存储所有的训练样本。
- **最近邻法** 找出与带识别样本最近邻的训练样本，并将该训练样本所属类别作为待识别测试样本的类别。简单实用，但是计算和存储代价高，且容易受噪声污染。

折衷方法：

- **引入集群操作** 把属于同一类的样本集合划分为若干个有凝聚性的子集（要求大于一定数量），取每个子集平均样本作为标准样本。然后用最近邻法分析。
- **K近邻法** 把待测试样本X到各个训练样本的距离都算出来，排序，取最近的K个距离，到哪个集合的最多的。（比最近邻法取了更多的样本信息）

2.3 线性可分情况下的几何分类法

通过几何方法把空间分成对应于不同区域的方法称为几何分类法。本节讨论集合分类器设计的一般方法。若采取欧氏距离作为距离测量，则判别函数具有线性函数的性质。

- **决策面函数** 函数 $G(X)$ 可以作为一个决策面函数， $G(X)=0$ 确定的曲面两侧属于不同的类别。一种简单的方式是各坐标的线性组合。权向量和决策面正交。 $G(X)$ 表示了点X到决策面的距离。

在多类情况下：

- 总体线性可分：使用一个超平面将任意一个类别的样本同其他所有类别的样本分开
- 成对线性可分：使用一个超平面将任意两个类别的样本分开

总体线性可分下的判决：直接选择某一类

成对线性可分下的判决：涉及该模式的所有决策面函数均为正。

成对线性可分且没有不确定区域：每个类别定义一个判别函数，采取最大值法。各判定面函数不独立，独立决策面个数是N-1（N为类别数量）。成对线性可分问题不一定有解。

线性判别函数的参数确定

对已知分类的每个向量X，每个求出 $X_k^TW=0$的平面，其解必为在空间中一个过原点的平面，组成的锥形即为对应的解区。解区内任何一点都有资格叫做解权向量W.

求解一般步骤：

1. 采集样本，求出已有的集合
2. 确定准则函数 $J=J(W,X)$，W和X一般为加长型向量
3. 求准则函数的极值获得解权向量W

准则函数J的一种算法是感知器算法： $J_p(W,X)=\sum\limits_{X\in X_W}-W^TX$，其中 X_W 指所有被W错分的样本。对某个W若存在分类错误则用 ρ 乘以X进行增减。

另一种是梯度下降法，对每个W寻找J在W处的梯度，根据步长决定下一个W在哪。

定理2.1 收敛性定理：若线性可分，则以上算法一定能得到正确解。

还有一种是最小平方误差法，设

$$[X]=\begin{bmatrix}X_1^T\\X_2^T\\\cdots\\X_N^T\end{bmatrix}$$

则 $W=[[X]^TX]^{-1}[X]^Tb=X^{\#}b$， $[X]^{\#}$ 称为X的伪逆。误差向量满足 $e=[X]W-b=([X]\,[X]^{\#}-I)b$，一般不为零。

一种新的迭代算法：LMSE，初始化 $W(0)=[X]^{\#}b(0)$，迭代为 $e(k)=[X]W(k)-b(k)$，当e(k)接近0时结束，有正有负则继续进行； $W(k+1)=W(k)+\rho[X]^{\#}(e(k)+|e(k)|)$，且 $b(k+1)=b(k)+\rho(e(k)+|e(k)|)$，若e(k)分量全为负则线性不可分。

2.4 非线性可分情况下的几何分类法

单超平面法：先画出一个只含一类的部分，剩下部分逐个尝试，构成折线分割的超平面。

双超平面法：用一对平行线划成三部分，使得有两个部分分别只含一类，递归进行。

位势法：就像空间分布正负电荷，在电势0点可以视为分界面。

2.5 线性可分问题的非迭代解法

Fisher解法：投影， $y_k=W^TX_k$， $m_x^l=\frac{1}{nl}\sum\limits_{X_k\in\omega l}X_k$， $m_y^l=\frac{1}{nl}\sum\limits_{X_k\in\omega l}y_k=W^Tm_x^l$

定义 $S_b=(m_x^i-m_x^j)(m_x^i-m_x^j)^T$ 为类间离散程度， $S_l=\sum\limits_{y_k\in\omega l}(y_k-m_y^l)^2$ 为类内离散度，则 $S_i^2+S_j^2=W^TS_wW$，其中 S_w 称为类内总离散度矩阵

$$S_w=\left[\sum_{X_k\in\omega i}(X_k-m_x^i)(X_k-m_x^i)^T+\sum_{X_k\in\omega j}(X_k-m_x^j)(X_k-m_x^j)^T\right]$$

投影方向 $W^*=S_w^{-1}(m_x^i-m_x^j)$， 称为Fisher线性判别式。

2.6 最优分类超平面

寻找一个与样本点间距最大的分类器。

推导看不懂，P70，直接给结论：若 N_{SV} 为训练样本中支持向量的个数， 最优权向量 W^* 为 $W^*=\sum\limits_{X_k\in SV}\lambda_ky_kX_k$，其中 λ_k 为待定拉格朗日系数，则

$$w_{d+1}=\frac{1}{N_{SV}}\left(\sum_{X_k\in SV}y_k-W^{*T}\sum_{X_k\in SV}X_k\right)$$