中国人民大學

RENMIN UNIVERSITY OF CHINA

XXXX

XXXXX

学院 ________________________________

专业 ________________________________

学号 ________________________________

姓名 ________________________________

2025 年 8 月 28 日

摘 要

XXXXX

关键词：XXX

目录

1　问题设定与基本概念

给定 n 个样本，特征维度为 d。每个样本 $X_i \in \mathbb{R}^d$，部分属于类别 C，其余不属于 C。将样本视为 $\mathbb{R}^d$ 中的点。

判别函数 / 预测函数 / 判别式（discriminant function）定义为标量函数 $f(x)$，满足

$$f(x) > 0 \Rightarrow x \in C, \qquad f(x) \leq 0 \Rightarrow x \notin C.$$

决策边界（decision boundary）定义为零水平集

$$\{x \in \mathbb{R}^d : f(x) = 0\}.$$

它通常是 $\mathbb{R}^d$ 中的一个 $(d-1)$ 维曲面，即 f 在等值 0 处的**等值面**（isosurface）；类似地还有 $\{x : f(x) = 1\}$ 等其它等值面。过度贴合训练集而产生蜿蜒曲线将导致**过拟合**。

2　线性分类器与超平面

若取线性判别函数

$$f(x) = w \cdot x + \alpha,$$

则决策边界为

$$H = \{x : w \cdot x = -\alpha\},$$

称 H 为**超平面**（d=2 为直线，d=3 为平面）。超平面的三条核心性质：

1. 维度为 $d-1$，将 $\mathbb{R}^d$ 切分为两半；
2. 平直（由一次方程定义，无曲率）；
3. 无界（无限延伸）。

2.1　法向量与正交性质

若 $x, y \in H$，则

$$w \cdot (y - x) = 0.$$

因此 w 垂直于 H 上的任意方向，w 称为 H 的**法向量**。

2.2　带符号距离与原点距离

若 w 为单位向量，则 $f(x) = w \cdot x + \alpha$ 是点到超平面 H 的**带符号距离**，正负号由法向量一侧决定；H 到原点的距离为 $|\alpha|$，且 $\alpha = 0$ 当且仅当 H 过原点。若 w 非单位向量，可将 w, α 同时除以 $\|w\|$ 归一化。w 与 α 的系数统称为**权重 / 回归系数**（weights）。

2.3 点到超平面的距离（通式）

对任意 $w \neq 0$，点 x_0 到 H：$w \cdot x + \alpha = 0$ 的距离

$$d(x_0, H) = \frac{|w \cdot x_0 + \alpha|}{\|w\|}.$$

当 w 为单位向量时，$f(x)$ 即上述距离的带符号形式。

3 质心法（Centroid Method）

记正类与负类（非 C）的样本均值为

$$\mu_C = \frac{1}{|C|} \sum_{X_i \in C} X_i, \qquad \mu_X = \frac{1}{|X|} \sum_{X_i \notin C} X_i.$$

采用判别函数

$$f(x) = (\mu_C - \mu_X) \cdot x - (\mu_C - \mu_X) \cdot \frac{\mu_C + \mu_X}{2}.$$

几何意义：法向量为 $\mu_C - \mu_X$；决策边界为连接 μ_C 与 μ_X 线段的**垂直平分超平面**。在某些数据上（即便线性可分）此方法仍可能误分；但当正负类分别来自两个高斯分布且协方差矩阵相同时，常表现良好。亦可调节标量项 α（不改变法向量）以减少误分。

与贝叶斯最优分类器的联系 若 $x|C_k \sim \mathcal{N}(\mu_k, \sigma^2 I)$ 且先验相近，则

$$\log p(x|C_1) - \log p(x|C_2) \propto -\tfrac{1}{2\sigma^2}\left(\|x - \mu_1\|^2 - \|x - \mu_2\|^2\right),$$

等价于“判给更近的质心”。此时质心法与 LDA 的线性边界一致，达到贝叶斯最优；若协方差不同，则最优边界为二次曲面（QDA）。

4 线性可分与感知机（Perceptron）

线性可分：存在某个超平面能将全部训练点正确划分。

4.1 算法设定

样本行向量 X_i 存于矩阵 X 的第 i 行，标签

$$y_i = \begin{cases} 1, & X_i \in C, \\ -1, & X_i \notin C. \end{cases}$$

为简便先考虑过原点的边界（后续可加偏置）。

4.2 约束、损失与风险

目标是找到 w 使

$$y_i(X_i \cdot w) \geq 0 \quad (\forall i).$$

定义分段线性**损失函数**

$$L(z, y_i) = \begin{cases} 0, & y_i z \geq 0, \\ -\,y_i z, & y_i z < 0, \end{cases}$$

并令**风险函数**

$$R(w) = \frac{1}{n} \sum_{i=1}^{n} L(X_i \cdot w, y_i) = \frac{1}{n} \sum_{i:\, y_i X_i \cdot w < 0} \big(-\,y_i\, X_i \cdot w\big).$$

若 w 正确分类全部样本则 $R(w) = 0$；否则 $R(w) > 0$，训练目标为

$$\min_{w} R(w).$$

4.3 与梯度下降的关系与更新

对误分类样本 i，有 $\nabla_w L(X_i \cdot w, y_i) = -y_i X_i$。随机梯度下降更新为

$$w \leftarrow w - \eta \nabla_w L = w + \eta\, y_i X_i,$$

当学习率 $\eta = 1$ 即经典感知机更新。感知机针对线性可分数据“慢但正确”；若数据不可分则不收敛。它不追求最大间隔，鲁棒性不如 SVM。