模式识别笔记

陈鸿峥

2019.12*

目录

1 简介

机器学习侧重于处理的算法，而模式识别则包括了数据预处理、实际运算和数据输出的完整过程。

- 模式识别：涵盖的范围广，包括特征提取、特征选择、降维、各种分类器等。
- 机器学习：主要是讲学习，更多关于分类器如何训练模型，而不涉及特征方面的知识。

良好特征的四个特点：

- 可区别性（不同类）
- 可靠性（同类）
- 独立性（特征之间）
- 参数少（复杂性）

一个对象的所有特征参数组成特征向量。同样需要从高维测量空间（样本）中提取特征映射到低维特征空间。

模式识别分为两类

- 结构/句法模式识别
- 统计/神经网络模式识别

*Build 20191207

2 贝叶斯决策论

2.1 离散变量

处于类别ω_i并具有特征值x，有后验概率[1]

$$\mathbb{P}\left(\omega_i \mid x\right)=\frac{p(x \mid \omega_i)\mathbb{P}\left(\omega_i\right)}{p(x)}$$

即

$$posterior=\frac{likelihood \times prior}{evidence}$$

无论什么情况，当我们观察到特定的x，对于二分类问题有错误率

$$\mathbb{P}\left(error \mid x\right)=\begin{cases}\mathbb{P}\left(\omega_1 \mid x\right) & \text{决策}\omega_2 \\ \mathbb{P}\left(\omega_2 \mid x\right) & \text{决策}\omega_1\end{cases}=\min[\mathbb{P}\left(\omega_1 \mid x\right),\mathbb{P}\left(\omega_2 \mid x\right)]$$

平均错误概率可表示为

$$\mathbb{P}\left(error \mid x\right)=\int_{-\infty}^{\infty}\mathbb{P}\left(error, x\right)\,\mathrm{d}x=\int_{-\infty}^{\infty}\mathbb{P}\left(error \mid x\right)p(x)\,\mathrm{d}x$$

注意$p(x)$是证据，可以看为是固定分布（常量）。

定理 1 (贝叶斯决策/最小错误率准则)**.** *若$P(\omega_1 \mid x)>P(\omega_2 \mid x)$，则判定类别为$\omega_1$；否则判为$\omega_2$。依照这种准则可以获得最小错误率，即$P(error \mid x)=\min[P(\omega_1 \mid x),P(\omega_2 \mid x)]$*

2.2 连续变量

考虑特征向量$\mathbf{x}\in\mathbb{R}^d$（$\mathbb{R}^d$称为特征空间），令$\{\omega_1,\ldots,\omega_c\}$表示有限的$c$个类别集，$\{\alpha_1,\ldots,\alpha_a\}$表示有限的$a$种可能采取的行为集，损失函数(loss)$\lambda(\alpha_t \mid \omega_j)$描述类别状态为$\omega_j$时采取行动$\alpha_i$的风险。 $p(\mathbf{x} \mid \omega_j)$表示在真实类别为$\omega_j$的条件下$\mathbf{x}$的概率密度函数，$P(\omega_j)$表示类别处于状态$\omega_j$时的先验概率，后验概率$P(\omega_j \mid \mathbf{x})$则通过贝叶斯公式

$$P(\omega_j \mid \mathbf{x}=\frac{p(\mathbf{x} \mid \omega_j)P(\omega_j)}{p(\mathbf{x})})$$

计算得到，证据变为

$$p(\mathbf{x})=\sum_{j=1}^{c}p(\mathbf{x} \mid \omega_j)P(\omega_j)$$

与行动α_i相关联的风险(risk)为

$$R(\alpha_i \mid \mathbf{x})=\sum_{j=1}^{c}\lambda(\alpha_i \mid \omega_j)\mathbb{P}\left(\omega_j \mid \mathbf{x}\right)$$

[1]通常用$p(\cdot)$代表概率密度函数（连续变量），用$\mathbb{P}\left(\cdot\right)$代表概率质量函数（离散变量）

进而得到总损失

$$R=\int R(\alpha(\mathbf{x})\mid\mathbf{x})p(\mathbf{x})\ \mathrm{d}\mathbf{x}$$

因此得到连续情形下的贝叶斯决策论：

定理 2. 为最小化R，计算条件概率

$$R(\alpha_i\mid\mathbf{x})=\sum_{j=1}^{c}\lambda(\alpha_i\mid\omega_j)\mathbb{P}\left(\omega_j\mid\mathbf{x}\right),\ \forall i=1,\ldots,a$$

选择α_i使得$R(\alpha_i\mid\mathbf{x})$最小，进而最小化总的风险即称为贝叶斯风险，记为$R^*$

对称损失/0-1损失

$$\lambda(\alpha_i\mid\omega_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}\qquad i,j=1,2,\ldots,c$$

有条件风险

$$\begin{cases}R(\alpha_1\mid\mathbf{x}) & =\lambda_{11}P(\omega_1\mid\mathbf{x})+\lambda_{12}P(\omega_2\mid\mathbf{x})\\ R(\alpha_2\mid\mathbf{x}) & =\lambda_{21}P(\omega_1\mid\mathbf{x})+\lambda_{22}P(\omega_2\mid\mathbf{x})\end{cases}$$