UCAS 模式识别-笔记 1-贝叶斯决策

一、统计模式识别分类

1.生成模型：基于概率密度的方法，利用参数估计或非参数估计求解概率密度，从而得到后验概率进行决策（基于贝叶斯决策理论）。不同类别分别学习。目标是得到每个类别的分布函数。

参数方法：假设数据的概率密度函数形式已知，根据训练集估计模型的参数。

代表：HMM/贝叶斯网络

非参数方法：不知道概率密度形式，直接根据数据估计概率密度函数。

代表：KNN/Parzen 窗

半参数法：GM，混合高斯知道参数的表达形式，但是里面又有隐变量，不明确知道。

2.判别模型：没有假设概率密度函数形式，直接根据数据集学习后验概率、判别函数、决策面，得到的是判别信息。所有类别样本一起学习，找到类别之间的区别。目标是分类误差最小。

代表：神经网络、svm、决策树、boosting

3.混合模型： 生成模型 ＋ 判别学习

二、贝叶斯决策理论

1.最小错误率准则——最大后验概率准则

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

哪个类别的后验概率大，就判为哪个类。

2.最小风险准则——贝叶斯决策的一般形式，引入了代价函数

代价函数：真实类别为 w_j，判为第α_i类的代价。

$$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$$

条件风险：x 判为第 i 类的总代价。注意这里，变量是α_i，x 已知，判为不同的类别，损失不一样。

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$$

最小风险决策：看 x 判为哪一类的风险（损失）最小，就判定为哪一类。

$$\arg\min_i R(\alpha_i \mid x)$$

3.两者的关系：当最小风险准则中，代价函数是 0-1 函数时，两者是等价的。值得注意的是，当先验概率相等时，最大后验概率准则又和最大似然准则等价。

下面，就以两类分类的例子解释以上结论：

代价函数为 0-1 损失：

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

条件风险展开求和如下：

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x}) \\ &= \sum_{j\neq i} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

所以，$R(\alpha_1|x) = 1 - P(w_1|x), R(\alpha_2|x) = 1 - P(w_2|x)$.

根据最小风险决策是找到使得条件风险最小的 i，即$R(\alpha_1|x)< R(\alpha_2|x)$，则判为$w_1$。化简可得此时 $P(w_1|x) > P(w_2|x)$，这恰好是最大后验概率准则下的判决。

由此证明当代价函数为 0-1 损失时，两个决策原则等价。

进一步，我们根据贝叶斯公式对$P(w_1|x) > P(w_2|x)$展开可得$P(x|w_1)P(w_1) > P(x|w_2)P(w_2)$

当先验概率相等时，得到$P(x|w_1) > P(x|w_2)$，即和最大似然准则等价。

4.推广：带拒识的决策

动机：在有的情况下，如果识别错误，代价会很大。这个时候，我们可以通过拒绝识别后，再交给人去处理，降低代价。

代价函数引入新的一类：

$$\lambda(\alpha_i \mid \omega_j)=\begin{cases} 0, & i=j \\ \lambda_s, & i \neq j \\ \lambda_r, & \text{reject} \end{cases}$$

求条件风险：

$$R_i(\mathbf{x})=\begin{cases} \lambda_s[1-P(\omega_i \mid \mathbf{x})], & i=1,\ldots,c \\ \lambda_r, & \text{reject} \end{cases}$$

根据最小风险准则决策：当拒识的风险比判错的风险小时，我就不判决了，直接拒识。否则我们就像之前一样找到 i 使得条件风险最小。

$$\arg\min_i R_i(\mathbf{x})=\begin{cases} \arg\max_i P(\omega_i \mid \mathbf{x}), & \text{if } \max_i P(\omega_i \mid \mathbf{x}) > 1-\lambda_r/\lambda_s \\ \text{reject}, & \text{otherwise} \end{cases}$$

三、判别函数和决策面——分类器的表示

1.定义

判别函数：表征每一类的广义似然度，哪个类的似然度越大，就判为哪一类。比如可以用后验概率。

决策面：两类判别函数相等的点的集合

2.高斯密度下的判别函数、决策面：如果协方差矩阵$\Sigma_i = \sigma^2 I$或者相等$\Sigma_i = \Sigma$，则它的判别函数是线性判别函数(LDF)，如果协方差矩阵是任意的，则是二次判别函数(QDF)。

证明如下：

这里，我们采用 likehood 作为判别函数：

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

其中，

$$p(\mathbf{x} \mid \omega_i)=\frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right]$$

根据协方差分为以下三种情况：

- Case 1: $\Sigma_i=\sigma^2 I$

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

– Euclidean distance $\|\mathbf{x}-\boldsymbol{\mu}_i\|^2 = (\mathbf{x}-\boldsymbol{\mu}_i)^t(\mathbf{x}-\boldsymbol{\mu}_i)$

– 展开二次式 $(\mathbf{x}-\boldsymbol{\mu}_i)^t(\mathbf{x}-\boldsymbol{\mu}_i)$

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}[\mathbf{x}^t\mathbf{x} - 2\boldsymbol{\mu}_i^t\mathbf{x} + \boldsymbol{\mu}_i^t\boldsymbol{\mu}_i] + \ln P(\omega_i)$$

– 忽略与类别无关项，得到线性判别函数

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0}$$

$$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i \qquad w_{i0} = \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i^t\boldsymbol{\mu}_i + \ln P(\omega_i)$$

– 二类决策面(判别函数相等的点构成)

$$g_i(\mathbf{x}) = g_j(\mathbf{x})$$

$$\Rightarrow \mathbf{w}^t(\mathbf{x}-\mathbf{x}_0) = 0 \qquad \mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

- Case 2: $\Sigma_i = \Sigma$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

$$\Rightarrow g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i)$$

– 展开二次式 $(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$

线性判别函数！ $g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0}$

$$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i \qquad w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i)$$

– 二类决策面 $g_i(\mathbf{x}) = g_j(\mathbf{x})$

$$\Rightarrow \mathbf{w}^t(\mathbf{x}-\mathbf{x}_0) = 0 \qquad \mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\ln[P(\omega_i)/P(\omega_j)]}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

  - 注意跟μ_1-μ_2的关系，决策面不一定与之垂直
  - 当$P(\omega_1) = P(\omega_2)$，决策面经过$(\mu_1+\mu_2)/2$

- Case 3: Σ_i= arbitrary

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

$$g_i(\mathbf{x}) = \mathbf{x}^t\mathbf{W}_i\mathbf{x} + \mathbf{w}_i^t\mathbf{x} + w_{i0}$$

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1} \qquad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

由此可以看到，当属于 case1/case2 时，判别函数为 LDF，case3 中，判别函数为 QDF。