1. 贝叶斯判定准则

2. 多元正态分布参数的极大似然估计

3. 朴素贝叶斯分类器

贝叶斯判定准则：为最小化总体风险，只需在每个样本上选择那个能使条件风险 $R(c|x)$ 最小的类别标记，即

$$h^*(x) = \arg\min_{c \in y} R(c|x)$$

此时，h^* 称为贝叶斯最优分类器。

已知条件风险 $R(c|x)$ 的计算公式为

$$R(c_i|x) = \sum_{j=1}^{N} \lambda_{ij} P(c_j|x) \quad \text{书上式7.1}$$

若目标是最小化分类错误率，则误判损失 λ_{ij} 对应为 0/1 损失，也即

$$\lambda_{ij} = \begin{cases} 0, & if\ i=j \\ 1, & otherwise \end{cases}$$

此时条件风险 $R(c|x) = 1 - P(c|x)$

$$R(c_i|x) = 1 \times P(c_1|x) + \dots + 1 \times P(c_{i-1}|x)$$
$$+ 0 \times P(c_i|x) + 1 \times P(c_{i+1}|x) + \dots + 1 \times P(c_N|x)$$
$$= P(c_1|x) + \dots + P(c_{i-1}|x) + P(c_{i+1}|x) + \dots + P(c_N|x)$$

由于 $\sum_{j=1}^{N} P(c_j|x) = 1$，所以 $R(c_i|x) = 1 - P(c_i|x)$ 　此式即为7.5

于是，最小化错误率的贝叶斯最优分类器为

$$h^*(x) = \arg\min_{c \in y} R(c|x) = \arg\min_{c \in y} (1 - P(c|x)) = \arg\max_{c \in y} P(c|x)$$

多元正态分布参数的极大似然估计

已知对数似然函数为

$$LL(\theta_c) = \sum_{x \in D_c} \log P(x|\theta_c) \quad \text{书上式7.10}$$

为了便于后续计算，我们令log的底数为e，则对数似然函数可化为

$$LL(\theta_c) = \sum_{x \in D_c} \ln P(x|\theta_c)$$

由于 $P(x|\theta_c) = P(x|c) \sim N(\mu_c, \sigma_c^2)$，那么

$$P(x|\theta_c) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_c|}} \exp\left(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)\right)$$

其中，d 表示 x 的维数，$\Sigma_c = \sigma_c^2$ 为对称正定协方差矩阵，$|\Sigma_c|$ 表示 Σ_c 的行列式，将上式代入对数似然函数可得

$$LL(\theta_c) = \sum_{x \in D_c} \ln\left[\frac{1}{\sqrt{(2\pi)^d |\Sigma_c|}} \exp\left(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)\right)\right]$$

令 $|D_c| = N$，则对数似然函数可化为

$$LL(\theta_c) = \sum_{i=1}^{N} \ln\left[\frac{1}{\sqrt{(2\pi)^d |\Sigma_c|}} \exp\left(-\frac{1}{2}(x_i-\mu_c)^T \Sigma_c^{-1} (x_i-\mu_c)\right)\right]$$

$$= \sum_{i=1}^{N} \ln\left[\frac{1}{\sqrt{(2\pi)^d}} \cdot \frac{1}{\sqrt{|\Sigma_c|}} \cdot \exp\left(-\frac{1}{2}(x_i-\mu_c)^T \Sigma_c^{-1} (x_i-\mu_c)\right)\right]$$

如常被用来学习高斯混合模型(Gaussian mixture model, 简称 GMM) 的参数;
9.4 节将介绍的 k 均值聚类算法…
的分析…

③

$$= \sum_{i=1}^{N}\left\{\ln\frac{1}{\sqrt{(2\pi)^d}}+\ln\frac{1}{\sqrt{|\Sigma_c|}}+\ln\left[\exp\left(-\frac{1}{2}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right)\right]\right\}$$

$$LL(\theta_c)=\sum_{i=1}^{N}\left\{-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_c|-\frac{1}{2}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right\}$$

$$=-\frac{Nd}{2}\ln 2\pi-\frac{N}{2}\ln|\Sigma_c|-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)$$

则参数 θ_c 的极大似然估计 $\hat{\theta}_c$ 为 $\hat{\theta}_c=\arg_{\theta_c}\max LL(\theta_c)$
所以接下来只需求出使得对数似然函数 $LL(\theta_c)$ 取到最大值的 $\hat{\mu}_c$ 和 $\hat{\Sigma}_c$，也就求出了 $\hat{\theta}_c$.

对 $LL(\theta_c)$ 关于 μ_c 求偏导

$$\frac{\partial LL(\theta_c)}{\partial \mu_c}=\frac{\partial}{\partial \mu_c}\left[-\frac{Nd}{2}\ln(2\pi)-\frac{N}{2}\ln|\Sigma_c|-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

$$=\frac{\partial}{\partial \mu_c}\left[-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

$$=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu_c}\left[(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

$$=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu_c}\left[(x_i^T-\mu_c^T)(\Sigma_c^{-1}x_i-\Sigma_c^{-1}\mu_c)\right]$$

$$=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu_c}\left[x_i^T\Sigma_c^{-1}x_i-x_i^T\Sigma_c^{-1}\mu_c-\mu_c^T\Sigma_c^{-1}x_i+\mu_c^T\Sigma_c^{-1}\mu_c\right]$$

由于 $x_i^T\Sigma_c^{-1}\mu_c$ 的计算结果为标量，所以

$$x_i^T\Sigma_c^{-1}\mu_c=(x_i^T\Sigma_c^{-1}\mu_c)^T=\mu_c^T(\Sigma_c^{-1})^Tx_i=\mu_c^T(\Sigma_c^T)^{-1}x_i=\mu_c^T\Sigma_c^{-1}x_i$$

$$\frac{\partial LL(\theta)}{\partial \mu_c}=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu_c}\left[x_i^T\Sigma_c^{-1}x_i-2x_i^T\Sigma_c^{-1}\mu_c+\mu_c^T\Sigma_c^{-1}\mu_c\right]$$

④

由矩阵微分公式 $\frac{\partial a^Tx}{\partial x}=a$ $\quad\frac{\partial x^TBx}{\partial x}=(B+B^T)x$ 可得

$$\frac{\partial LL(\theta_c)}{\partial \mu_c}=-\frac{1}{2}\sum_{i=1}^{N}\left[0-(2x_i^T\Sigma_c^{-1})^T+(\Sigma_c^{-1}+(\Sigma_c^{-1})^T)\mu_c\right]$$

$$=-\frac{1}{2}\sum_{i=1}^{N}\left[-2\Sigma_c^{-1}x_i+2\Sigma_c^{-1}\mu_c\right]$$

$$=\sum_{i=1}^{N}\Sigma_c^{-1}x_i-N\Sigma_c^{-1}\mu_c$$

令偏导数等于0可得

$$\frac{\partial LL(\theta_c)}{\partial \mu_c}=\sum_{i=1}^{N}\Sigma_c^{-1}x_i-N\Sigma_c^{-1}\mu_c=0$$

$$N\Sigma_c^{-1}\mu_c=\sum_{i=1}^{N}\Sigma_c^{-1}x_i$$

$$N\Sigma_c^{-1}\mu_c=\Sigma_c^{-1}\sum_{i=1}^{N}x_i$$

$$N\mu_c=\sum_{i=1}^{N}x_i$$

$$\mu_c=\frac{1}{N}\sum_{i=1}^{N}x_i\Rightarrow\hat{\mu}_c=\frac{1}{N}\sum_{i=1}^{N}x_i$$

此式即为 7.12.

对 $LL(\theta_c)$ 关于 Σ_c 求偏导

$$\frac{\partial LL(\theta_c)}{\partial \Sigma_c}=\frac{\partial}{\partial \Sigma_c}\left[-\frac{Nd}{2}\ln(2\pi)-\frac{N}{2}\ln|\Sigma_c|-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

$$=\frac{\partial}{\partial \Sigma_c}\left[-\frac{N}{2}\ln|\Sigma_c|-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

$$=-\frac{N}{2}\cdot\frac{\partial}{\partial \Sigma_c}\left[\ln|\Sigma_c|\right]-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \Sigma_c}\left[(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]$$

由矩阵微分公式 $\frac{\partial|X|}{\partial X}=|X|(X^{-1})^T$， $\frac{\partial a^TX^{-1}b}{\partial X}=-X^{-T}ab^TX^{-T}$

NO. 5

DATE

$$\frac{\partial LL(\theta_c)}{\partial \Sigma_c} = -\frac{N}{2}\cdot\frac{1}{|\Sigma_c|}\cdot|\Sigma_c|\cdot(\Sigma_c^{-1})^T - \frac{1}{2}\sum_{i=1}^{N}\left[-\Sigma_c^{-T}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right]$$

$$= -\frac{N}{2}\cdot(\Sigma_c^{-1})^T - \frac{1}{2}\sum_{i=1}^{N}\left[-\Sigma_c^{-T}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right]$$

$$= -\frac{N}{2}\Sigma_c^{-1} + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma_c^{-1}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right]$$

令偏导数等于0.

$$\frac{\partial LL(\theta_c)}{\partial \Sigma_c} = -\frac{N}{2}\Sigma_c^{-1} + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma_c^{-1}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right] = 0$$

$$-\frac{N}{2}\Sigma_c^{-1} = -\frac{1}{2}\sum_{i=1}^{N}\left[\Sigma_c^{-1}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right]$$

$$N\Sigma_c^{-1} = \sum_{i=1}^{N}\left[\Sigma_c^{-1}(x_i-u_c)(x_i-u_c)^T\Sigma_c^{-1}\right]$$

$$N\Sigma_c^{-1} = \Sigma_c^{-1}\left[\sum_{i=1}^{N}(x_i-u_c)(x_i-u_c)^T\right]\Sigma_c^{-1}$$

$$N = \Sigma_c^{-1}\left[\sum_{i=1}^{N}(x_i-u_c)(x_i-u_c)^T\right]$$

$$\Sigma_c = \frac{1}{N}\sum_{i=1}^{N}(x_i-u_c)(x_i-u_c)^T \Rightarrow \hat{\Sigma}_c = \frac{1}{N}\sum_{i=1}^{N}(x_i-u_c)(x_i-u_c)^T$$

此式即为7.13

朴素贝叶斯分类器

NO. 6

DATE

已知最小化分类错误率的贝叶斯最优分类器为

$$h^*(x) = \arg\max_{c\in y} P(c|x)$$

又由贝叶斯定理可知 $P(c|x) = \frac{P(x,c)}{P(x)} = \frac{P(c)P(x|c)}{P(x)}$

所以 $h^*(x) = \arg\max_{c\in y}\frac{P(c)P(x|c)}{P(x)} = \arg\max_{c\in y} P(c)P(x|c)$

已知属性条件独立性假设为 $P(x|c) = P(x_1,x_2,\dots,x_d|c) = \prod_{i=1}^{d}P(x_i|c)$

$$h^*(x) = \arg\max_{c\in y} P(c)\cdot\prod_{i=1}^{d}P(x_i|c)$$