概率分类法

一、生成模型与判别模型(Generative Model and Discriminative Model)

① 生成模型

设特征为 X，类别为 y，生成模型估计的是概率 $p(X,y)$，用到数据为 (X_1,y_1)，$(X_2,y_2)\ldots(X_N,y_N)$

② 判别模型

判别模型估计的是 $p(y|X)$

例子：语言识别、识别猫和狗

③ 典型的生成模型：Gaussian，Mixture of Gaussi… Hidden Markov Models，Markov Random Fields

典型的判别模型：SVM，Neural Networks(看情况)，Nearest Neighbor，Conditional Random Field.

二、概率密度估计

① Gaussian (生成模型)

设特征为 X，类别 $y=1$ 与 $y=-1$。设

$$p(X|y=1)=\frac{1}{\sqrt{(2\pi)^d|\Sigma_1|}}\exp\left\{-\frac{1}{2}(X-\mu_1)^T\Sigma_1^{-1}(X-\mu_1)\right\}$$

$$p(X|y=-1)=\frac{1}{\sqrt{(2\pi)^d|\Sigma_{-1}|}}\exp\left\{-\frac{1}{2}(X-\mu_{-1})^T\Sigma_{-1}^{-1}(X-\mu_{-1})\right\}$$

注：一般来说，构造生成模型时，我们假设

$$p(X|y)=p(X|y,\theta)$$

，其中 θ 为待确定的参数，通过训练样本确定。在这个例子中，$\theta=\{\mu,\Sigma\}$

估计方法：最大似然(Maximum Likelihood)

设有训练样本：(X_1,y_1) $(X_2,y_2)\ldots(X_N,y_N)$，以估计 μ_1,Σ_1 为例。

ⓐ 挑选 $x_1, x_2 \ldots x_M$ 都属于 $y_i=1$ 那一类 $(i=1 \sim M)$
则有:

$$p(x_i|y_i=1) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_1|}} \exp\left\{-\frac{1}{2}(x_i-\mu_1)^T \Sigma_1^{-1}(x_i-\mu_1)\right.$$

$(i=1 \sim M)$

ⓑ 求 μ_1, Σ_1，使

$$E(\mu_1, \Sigma_1) = \sum_{i=1}^{M} \ln p(x_i|y_i=1)$$

最大。(最大似然估计)
在这一特例中，

$$E(\mu_1, \Sigma_1) = -\frac{M}{2}\ln|\Sigma_1| - \frac{1}{2}\sum_{i=1}^{M}(x_i-\mu_1)^T\Sigma_1^{-1}(x_i-\mu_1) + \text{常数}$$

因此:

$$\frac{\partial E}{\partial \mu_1} = 0 \Rightarrow \mu_1 = \frac{1}{M}\sum_{i=1}^{M} x_i$$

$$\frac{\partial E}{\partial (\Sigma^{-1})} = 0 \Rightarrow \frac{M}{2}\Sigma_1^T - \frac{1}{2}\sum_{i=1}^{M}(x_i-\mu_1)(x_i-\mu_1)^T = 0$$

$$\Rightarrow \Sigma_1 = \frac{1}{M}\sum_{i=1}^{M}(x_i-\mu_1)(x_i-\mu_1)^T$$

② Gaussian Mixture Model (混合高斯模型)

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$

其中:

$$\begin{cases} N(x|\mu_k, \Sigma_k) = \dfrac{1}{\sqrt{(2\pi)^d|\Sigma_k|}} \exp\left\{-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)\right. \\ \sum_{k=1}^{K} \pi_k = 1 \end{cases}$$

已知 N个X值$(x_1, x_2, \ldots x_N)$，需要求最大似然

$$E(\{\pi_k, \mu_k, \Sigma_k\}_{k=1\sim K})$$

$$= \sum_{i=1}^{N} \ln p(x_i)$$

$$= \sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d|\Sigma_k|}} \exp\left\{-\frac{1}{2}(x_i-\mu_k)^T \Sigma^{-1}(x_i-\mu_k)\right\}\right)$$

此式难以用求导法直接获得 $\{\pi_k, \mu_k, \Sigma_k\}_{k=1\sim K}$ 的解析表达式，需要用近似优化算法求解。

最常用的是EM算法(Expectation Maximization)算法。(参见讲义EM算法)，描绘如下。

Step1: 随机选取 $\{\pi_k, \mu_k, \Sigma_k\}_{k=1\sim K}$。

Step 2: (E step)

求 $$\overset{\gamma_{nk}}{\cancel{p(z_{nk})}} = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k)}$$

Step 3: (M Step)

$$\begin{cases} \mu_k^{new} = \dfrac{1}{N_k} \sum_{n=1}^{N} \overset{\gamma_{nk}}{\cancel{p(z_{nk})}}\, x_n \\ \Sigma_k^{new} = \dfrac{1}{N_k} \sum_{n=1}^{N} \overset{\gamma_{nk}}{\cancel{p(z_{nk})}}\, (x_n-\mu_k^{new})(x_n-\mu_k^{new})^T \\ \pi_k^{new} = \dfrac{N_k}{N} \end{cases}$$

其中 $$N_k = \sum_{n=1}^{N} \overset{\gamma_{nk}}{\cancel{p(z_{nk})}}$$

Step 4: 计算 $E(\{\pi_k^{new}, \mu_k^{new}, \Sigma_k^{new}\})$

$E(\{\pi_k, \mu_k, \Sigma_k\}_{k=1\sim K})$ 比较，重复 Step2, Step 3 直至 E 收敛。

③ 朴素贝叶斯分类器 (Naive Bayesian Classifier)

Input: ① a document d

② a fixed set of classes

$C = \{c_1, c_2, \dots c_N\}$

output: a predicted class $d \in c_j$ $(j = 1 \sim N)$

Traning set:

$(d_1\ c_1)\quad (d_2, c_2)\ \cdots\ (d_n, c_n)$

$$p(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

$$C_{map} = \arg\max_{c \in C} P(d|c)P(c)$$

$$= \arg\max_{c \in C} P(w_1, w_2, \dots w_n | c)P(c)$$

朴素贝叶斯假设，特征之间相互独立！即

$$P(w_1, w_2 \dots w_n | c) = P(w_1|c)P(w_2|c)\cdots P(w_n|c)$$

$$= \prod_{i=1}^{n} P(w_i|c)$$

$$C_{map} = \arg\max_{c_j \in C} P(c_j) \prod_{i=position}^{n} P(x_i|c_j)$$

$$P(C_j) = \frac{\text{Num of Documents belonging to } C_j}{\text{Num of all documents}}$$

$$P(w_i|C_j) = \frac{count(w_i, C_j)}{\sum_{w \in V} count(w, C_j)}$$

改进

$$P(w_i|C) = \frac{count(w_i, C) + 1}{\sum_{w \in V} count(w, C) + |V|}$$