Homework1

2019 年 11 月 7 日

1 1. Generative and Discriminative classifiers: Gaussian Bayes and Logistic Regression

1.1 Specific Gaussian naive Bayes classifiers and logistic regression

已知 y 是一个布尔型变量,它服从伯努利分布,即 $\pi = P(y=1)$ 以及 $P(Y=0) = 1-\pi$。对于一个二分类问题，Logistic regression 恰好合适。它的输出在 $(0,1)$ 区间内，大于某一阈值判断为正，反之则为负。

在这种情况下，他们二者的关系有如下推导:

$$\begin{aligned} P(y=1|\mathbf{x}) &= \frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x})} \\ &= \frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x}|y=0)P(y=0)+P(\mathbf{x}|y=1)P(y=1)} \end{aligned} \tag{1}$$

已知向量空间为 $\mathbf{x}$ 为 $\mathbf{x} = [x_1, \ldots, x_D]$，且对于每一个特征 x_i，它都符合如下高斯分布:

$\mathcal{N}(u_{ik}, \sigma_i)$，其中 σ_i 是高斯分布的标准差，它独立于 k。把这些条件带入1可得：

$$
\begin{aligned}
P(y=1|\mathbf{x}) &= \frac{\pi \prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_i)}{(1-\pi)\prod_{i=1}^{D} \mathcal{N}(u_{i0}, \sigma_i) + \pi \prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_i)} \\
&= \frac{1}{1 + \frac{1-\pi}{\pi} \frac{\prod_{i=1}^{D} \mathcal{N}(u_{i0}, \sigma_i)}{\prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_i)}} \\
&= \frac{1}{1 + \frac{1-\pi}{\pi} \frac{e^{-\sum_{i=1}^{D} \frac{(x_i-u_{i0})^2}{2\sigma_i^2}} \prod_{i=1}^{D}(\frac{1}{\sqrt{2\pi}\sigma_i})}{e^{-\sum_{i=1}^{D} \frac{(x_i-u_{i1})^2}{2\sigma_i^2}} \prod_{i=1}^{D}(\frac{1}{\sqrt{2\pi}\sigma_i})}} \\
&= \frac{1}{1 + \frac{1-\pi}{\pi} e^{\sum_{i=1}^{D} \frac{(x_i-u_{i1})^2-(x_i-u_{i0})^2}{2\sigma_i^2}}} \\
&= \frac{1}{1 + e^{(\ln\frac{1-\pi}{\pi} + \sum_{i=1}^{D}[\frac{(u_{i0}-u_{i1})x_i}{\sigma^2} + \frac{u_{i1}^2-u_{i0}^2}{2\sigma_i^2}])}} \\
&= \frac{1}{1 + e^{(w_0 + \sum_{i=1}^{D} w_i x_i)}}
\end{aligned}
\tag{2}
$$

1.2 General Gaussian naive Bayes classifiers and logistic regression

将 $\mathcal{N}(u_{ik}, \sigma_{ik})$ 带入1中可以得到下式：

$$
\begin{aligned}
P(y=1|\mathbf{x}) &= \frac{\pi \prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_{i1})}{(1-\pi)\prod_{i=1}^{D} \mathcal{N}(u_{i0}, \sigma_{i0}) + \pi \prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_{i1})} \\
&= \frac{1}{1 + \frac{1-\pi}{\pi} \frac{\prod_{i=1}^{D} \mathcal{N}(u_{i0}, \sigma_{i0})}{\prod_{i=1}^{D} \mathcal{N}(u_{i1}, \sigma_{i1})}} \\
&= \frac{1}{1 + \frac{1-\pi}{\pi} \frac{e^{-\sum_{i=1}^{D} \frac{(x_i-u_{i0})^2}{2\sigma_{i0}^2}} \prod_{i=1}^{D}(\frac{1}{\sqrt{2\pi}\sigma_{i0}})}{e^{-\sum_{i=1}^{D} \frac{(x_i-u_{i1})^2}{2\sigma_{i1}^2}} \prod_{i=1}^{D}(\frac{1}{\sqrt{2\pi}\sigma_{i1}})}} \\
&= \frac{1}{1 + e^{(\ln\frac{1-\pi}{\pi} + \sum_{i=1}^{D} \frac{\sigma_{i1}}{\sigma_{i0}} + \sum_{i=1}^{D}[\frac{(x_i-u_{i1})^2}{2\sigma_{i1}^2} - \frac{(x_i-u_{i0})^2}{2\sigma_{i0}^2}])}}
\end{aligned}
\tag{3}
$$

由于 σ_{i0} 和 σ_{i1} 不相同，因此上式不能化为2的形式。即 e 的指数无法化为 $w_o + \sum_{i=1}^{D} w_i x_i$ 的格式。

1.3 Gaussian Bayes classifiers and logistic regression

对于特征 $x_i, x_j (i \neq j)$ 不相互独立的情况下。有:

$$\begin{aligned} P(y=1|\mathbf{x}) &= \frac{1}{1+\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x}|y=1)P(y=1)}} \\ &= \frac{1}{1+\frac{1-\pi}{\pi}\frac{P(x_1,x_2|y=0)}{P(x_1,x_2|y=1)}} \end{aligned} \tag{4}$$

将 $P(x_1, x_2|y=k)$ 的公式带入4可得:

$$P(y=1|\mathbf{x}) = \frac{1}{1+\frac{1-\pi}{\pi}e^{M(x_1,x_2)}} \tag{5}$$

其中，$M(x_1, x_2)$ 为:

$$M(x_1,x_2) = \frac{N_1(x_1)+N_2(x_2)+N_{1,2}(x_1,x_2)}{2(1-\rho^2)\sigma_1^2\sigma_2^2}$$

其中:

$$N_1(x_1) = \sigma_2^2(u_{11} - u10)[u_{11} + u_{10} - 2x_1]$$

$$N_2(x_2) = \sigma_1^2(u_{21} - u20)[u_{21} - u_{20} - 2x_2]$$

$$N_{1,2}(x_1,x_2) = 2\rho\sigma_1\sigma_2[(u_{21} - u_{20})x_1 + (u_{11} - u_{10})x_2 + u_{10}u_{20} - u_{11}u_{21}]$$

显然，5中的变量 x_1, x_2 都是一次的，那么5式显然可以被整理为2的结果:

$$P(y=1|\mathbf{x}) = \frac{1}{1+e^{w_1x_1+w_2x_2+w_0}} \tag{6}$$