Pattern Recognition: Assignment 6

钟赟 202028013229148

2020 年 12 月 22 日

1 Generative and Discriminative classifiers: Gaussian Bayes and Logistic Regression

Recall that a generative classifier estimates $P(\mathbf{x}, y) = P(y)P(\mathbf{x} \mid y)$, while a discriminative classifier directly estimates $P(y \mid \mathbf{x})$.

1.1 Specific Gaussian naive Bayes classifiers and logistic regression

Consider a **specific class** of Gaussian naive Bayes classifiers where:

- y is a boolean variable following a Bernoulli distribution, with parameter $\pi = P(y = 1)$ and thus $P(Y = 0) = 1 - \pi$.
- $\mathbf{x} = [x_1, \ldots, x_D]^T$, with each feature x_i a continuous random variable. For each $x_i, P(x_i \mid y = k)$ is a Gaussian distribution $\mathcal{N}(\mu_{ik}, \sigma_i)$. Note that σ_i is the standard deviation of the Gaussian distribution, which does not depend on k.
- For all $i \neq j, x_i$ and x_j are conditionally independent given y (so called "naive" classifier).

Question: please show that the relationship between a discriminative classifier (say logistic regression) and the above specific class of Gaussian naive Bayes classifiers is precisely the form used by logistic regression.

Solution

$P(Y = 1|X) = \frac{P(Y=1)P(X|Y=1)}{P(Y=1)P(X|Y=1)+P(Y=0)P(X|Y=0)} = \frac{1}{1+\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}}$
分子分母取对数得：
$P(Y = 1|X) = \frac{1}{1+e^{\ln\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}}} = \frac{1}{1+e^{\ln\frac{P(Y=0)}{P(Y+1)}+\sum_i^D \ln\frac{P(X|Y=0)}{P(X|Y=1)}}}$
由题已知 $P(Y = 0) = 1 - \pi, P(Y = 1) = \pi$，故 $P(Y = 1|X) = \frac{1}{1+e^{\ln\frac{1-\pi}{\pi}+\sum_i^D \ln\frac{P(X|Y=0)}{P(X|Y=1)}}}$
其中

$$
\begin{aligned}
\sum_i^D \ln \frac{P(X|Y=0)}{P(X|Y=1)} &= \sum_i^D \ln \frac{\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\mu_{i0})^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\mu_{i1})^2}{2\sigma^2}}} \\
&= \sum_i^D \ln e^{-\frac{(x_i-\mu_{i0})^2-(x_i-\mu_{i1})^2}{2\sigma_i^2}} \\
&= \sum_i^D \frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2} \\
&= \sum_i^D \frac{X_i^2-2X_i\mu_{i1}+\mu_{i1}^2-X_i^2+2X_i\mu_{i0}-\mu_{i0}^2}{2\sigma_i^2} \\
&= \sum_i^D \frac{\mu_{i1}^2-\mu_{i0}^2+2X_i(\mu_{i0}-\mu_{i1})}{2\sigma_i^2} \\
&= \sum_i^D (\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2}+\frac{2X_i(\mu_{i0}-\mu_{i1})}{\sigma_i^2})
\end{aligned}
$$

因此 $P(Y=1|X)=\frac{1}{1+e^{\ln\frac{1-\pi}{\pi}+\sum_i^D(\ln\frac{X_i(\mu_{i0}-\mu_{i1})}{\sigma_i^2}+\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2})}}$，记为 $P(Y=1|X)=\frac{1}{1+e^{w_0+\sum_i^D w_iX_i}}$ ，其中 $w_i=\frac{X_i(\mu_{i0}-\mu_{i1})}{\sigma_i^2}$ ，则 $w_0=\ln\frac{1-\pi}{\pi}+\sum_i^D\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2}$

故 $P(Y=0|X)=1-P(Y=1|X)=\frac{e^{w_0+\sum_i^D w_iX_i}}{1+e^{w_0+\sum_i^D w_iX_i}}$

1.2 General Gaussian naive Bayes classifiers and logistic regression

Removing the assumption that the standard deviation σ_i of $P(x_i \mid y=k)$ does not depend on k. That is , for each x_i, $P(x_i \mid y=k)$ is a Gaussian distribution $\mathcal{N}(\mu_{ik},\sigma_{ik})$, where $i=1,\ldots,D$ and $k=0,1$.
Question: is the new form of $P(y \mid \mathrm{x})$ implied by this more general Gaussian naive Bayes classifier still the form used by logistic regression? Derive the new form of $P(y \mid \mathbf{x})$ to prove your answer.

Solution

根据题目 1.1 得， $P(Y=1|X)=\frac{1}{1+e^{\ln\frac{1-\pi}{\pi}+\sum_i^D\ln\frac{P(X|Y=0)}{P(X|Y=1)}}}$ ，其中 $P(X_i|y=k)\ \ N(\mu_{ik},\sigma_{ik})$

$$
\begin{aligned}
\sum_i^D \ln \frac{P(X|Y=0)}{P(X|Y=1)} &= \sum_i^D \ln e^{-\frac{(x_i-\mu_{i0})^2-(x_i-\mu_{i1})^2}{2\sigma_i^2}} \\
&= \sum_i^D \ln \frac{\sigma_{i1}}{\sigma_{i0}} e^{-(\frac{2\sigma_{i1}^2(X_i-\mu_{i0})^2-2\sigma_{i0}(x_i-\mu_{i1})^2)}{2\sigma_{i0}^2\sigma_{i1}^2}} \\
&= \sum_i^D \ln \frac{\sigma_{i1}}{\sigma_{i0}} e^{\frac{x_i^2(\sigma_{i0}^2-\sigma_{i1}^2)+2x_i(2\sigma_{i0}^2\mu_{i1}-2\sigma_{i1}^2\mu_{i0})+\sigma_{i0}^2\mu_{i1}^2-\sigma_{i1}^2\mu_{i0}^2}{2\sigma_{i0}^2\sigma_{i1}^2}}
\end{aligned}
$$

该式始终存在 x_i^2 无法消去，因此不满足逻辑回归的形式。

Gaussian Bayes classifiers and logistic regression

Now, consider the following assumptions for our Gaussian Bayes classifiers (without “naive”):

- y is a boolean variable following a Bernoulli distribution, with parameter $\pi = P(y=1)$ and thus $P(Y=0) = 1-\pi$.
- $\mathbf{x} = [x_1, x_2]^T$, i.e., we only consider two features for each sample, with each feature a continuous random variable. x_1 and x_2 are not conditional independent given y. We assume $P(x_1, x_2 \mid y=k)$ is a bivariate Gaussian distribution $\mathcal{N}(\mu_{1k}, \mu_{2k}, \sigma_1, \sigma_2, \rho)$, where μ_{1k} and μ_{2k} are means of x_1 and x_2, σ_1 and σ_2 are standard deviations of x_1 and x_2, and ρ is the correlation between x_1 and x_2. The density of the bivariate Gaussian distribution is:

$$P(x_1, x_2 \mid y=k) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\sigma_2^2(x_1-\mu_{1k})^2 + \sigma_1^2(x_2-\mu_{2k})^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_{1k})(x_2-\mu_{2k})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}\right]$$

Question: is the form of $P(y \mid \mathbf{x})$ implied by such not-so-naive Gaussian Bayes classifiers still the form used by logistic regression? Derive the form of $P(y \mid \mathbf{x})$ to prove your answer.

Solution

根据题目 1.1 得， $P(Y=1|X) = \frac{1}{1+e^{\ln\frac{1-\pi}{\pi}+\sum_i^D \ln\frac{P(X|Y=0)}{P(X|Y=1)}}}$

$$\sum_i^D \ln\frac{P(x_1x_2|Y=0)}{P(x_1x_2|Y=1)} = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}e^{1\frac{\sigma_2^2(x_i-\mu_{10})^2+\sigma_1^2(x_2-\mu_{20})^2-2\rho\sigma_1\sigma_2(x_1-\mu_{10})(x_2-\mu_{20})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}}}}{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}e^{1\frac{\sigma_2^2(x_i-\mu_{11})^2+\sigma_1^2(x_2-\mu_{21})^2-2\rho\sigma_1\sigma_2(x_1-\mu_{11})(x_2-\mu_{21})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}}}}$$

$$= e^{\frac{\sigma_2^2(x_i-\mu_{11})^2+\sigma_1^2(x_2-\mu_{21})^2-2\rho\sigma_1\sigma_2(x_1-\mu_{11})(x_2-\mu_{21})-\sigma_2^2(x_i-\mu_{10})^2+\sigma_1^2(x_2-\mu_{20})^2-2\rho\sigma_1\sigma_2(x_1-\mu_{10})(x_2-\mu_{20})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}}$$

$$= e^{\frac{\sigma^2(2x_1\mu_{10}-2x_1\mu_{11}+\mu_{11}^2-\mu_{10}^2)+\sigma_1^2(2x_2\mu_{20}-2x_2\mu_{21}+\mu_{21}^2-\mu_{20}^2)+2\rho\sigma_1\sigma_2(x_1(\mu_{21}-\mu_{20})+x_2(\mu_{11}-\mu_{10})-\mu_{11}\mu_{21}-\mu_{10}\mu_{20})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}}$$

其中 x_1^2, x_2^2, x_1x_2 都被消去了。

$$P(Y=1|X) = \frac{1}{1+e^{\ln\frac{1-\pi}{\pi}+\frac{x_1(2\sigma_2^2\mu_{10}-2\sigma_2^2\mu_{11}+2\rho\sigma_1\sigma_2\mu_{21})+x_2(2\sigma_1^2\mu_{20}-2\sigma_1^2\mu_{21}+2\rho\sigma_1\sigma_2\mu_{11})+\sigma_2^2\mu_{11}^2-\sigma_1^2\mu_{10}^2+\dots}{2(1-\rho^2)\sigma_1^2\sigma_2^2}}}$$

$$= \frac{1}{1+e^{w_0+w_1x_1+w_2x_2}}$$

其中

$$w_1 = \frac{2\sigma_2^2\mu_{10} - 2\sigma_2^2\mu_{11} + 2\rho\sigma_1\sigma_2\mu_{21}}{2(1-\rho^2)\sigma_1^2\sigma_2^2}$$

$\mathrm{w}_2 = \frac{2\sigma_2^2\mu_{20}-2\sigma_1^2\mu_{21}+2\rho\sigma_1\sigma_2\mu_{11}}{2(1-\rho^2)\sigma_1^2\sigma_2^2}$

$\mathrm{w}_0 =$
$ln\frac{1-\pi}{\pi} + \frac{\sigma_2^2\mu_{11}^2-\sigma_2^2\mu_{10}^2+\sigma_1^2\mu_{21}^2-\sigma_1^2\mu_{20}^2-2\rho\sigma_1\sigma_2(\mu_{20}+\mu_{10}+\mu_{11}\mu_{21}+\mu 10\mu_{20})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}$
满足逻辑回归的形式。同理 $P(Y=0|X) = 1-P(Y=1|X)$ 。