Assignment 1

1 Generative and Discriminative classifiers: Gaussian Bayes and Logistic Regression

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

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, $P(x_i|y = k)$ 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.

1.2 General Gaussian naive Bayes classifiers and logistic regression

Removing the assumption that the standard deviation σ_i of $P(x_i|y = k)$ does not depend on k. That is , for each x_i, $P(x_i|y = k)$ 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 $P(y|\mathbf{x})$ implied by this more general Gaussian naive Bayes classifier still the form used by logistic regression? Derive the new form of $P(y|\mathbf{x})$ to prove your answer.

1.3 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 $P(x_1, x_2|y = k)$ 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:

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

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

第六章必做作业答案

- **Specific Gaussian naive Bayes classifiers and logistic regression**

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.

Answer:

可知 logistic 回归的一般形式：

$$P(Y=1|X)=\frac{1}{1+\exp\left(w_0+\sum_{i=1}^{n} w_i x_i\right)}$$

$$P(Y=0|X)=\frac{\exp(w_0+\sum_{i=1}^{n} w_i x_i)}{1+\exp\left(w_0+\sum_{i=1}^{n} w_i x_i\right)}$$

根据上述特定高斯朴素贝叶斯分类器的假设，以及贝叶斯法则，有：

$$P(Y=1|X)=\frac{P(Y=1)P(X|Y=1)}{P(Y=1)P(X|Y=1)+P(Y=0)P(X|Y=0)}$$

$$=\frac{1}{1+\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}}=\frac{1}{1+\exp\left(\ln\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}\right)}$$

由给定 Y，x的条件独立性假设，可得：

$$P(Y=1|X)=\frac{1}{1+\exp\left(\ln\frac{P(Y=0)}{P(Y=1)}+\ln\left(\prod_i\frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)\right)}$$

$$=\frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi}+\sum_i\ln\frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)}$$

再根据$P(x_i|Y=y_k)$服从高斯分布$\mathcal{N}(\mu_{ik},\sigma_i)$，可得：

$$\sum_i\ln\frac{P(x_i|Y=0)}{P(x_i|Y=1)}=\sum_i\ln\frac{\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(x_i-\mu_{i0})^2}{2\sigma_i^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(x_i-\mu_{i1})^2}{2\sigma_i^2}\right)}$$

$$=\sum_i\ln\exp\left(\frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2}\right)$$

$$=\sum_i\frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2}$$

$$= \sum_i \frac{(x_i^2 - 2x_i\mu_{i1} + \mu_{i1}^2) - (x_i^2 - 2x_i\mu_{i0} + \mu_{i0}^2)}{2\sigma_i^2}$$

$$= \sum_i \frac{2x_i(\mu_{i0} - \mu_{i1}) + \mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2}$$

$$= \sum_i (\frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2} x_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2})$$

则：

$$P(Y=1|X) = \frac{1}{1 + \exp\left(\ln \frac{1-\pi}{\pi} + \sum_i (\frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2} x_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2})\right)}$$

等价于：

$$P(Y=1|X) = \frac{1}{1 + \exp\left(w_0 + \sum_{i=1}^n w_i x_i\right)}$$

其中

$$w_0 = \ln \frac{1-\pi}{\pi} + \sum_i \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2} \qquad w_i = \frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2}$$

满足 logistic 回归的一般形式，即证判别分类器（如 logistic 回归）与上述特定高斯朴素贝叶斯分类器之间的关系正是 logistic 回归所使用的形式。

- **General Gaussian naive Bayes classifiers and logistic regression**

Question: is the new form of $P(y|\mathbf{x})$ implied by this more general Gaussian naive Bayes classifier still the form used by logistic regression? Derive the new form of $P(y|\mathbf{x})$ to prove your answer.

Answer:

由上一小题的解答，可以到这一步：

$$P(Y=1|X) = \frac{1}{1 + \exp\left(\ln \frac{P(Y=0)}{P(Y=1)} + \ln\left(\prod_i \frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)\right)}$$

$$= \frac{1}{1 + \exp\left(\ln \frac{1-\pi}{\pi} + \sum_i \ln \frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)}$$

再根据$P(x_i|Y=y_k)$服从高斯分布$\mathcal{N}(\mu_{ik}, \sigma_{ik})$，可得：

$$\sum_i \ln \frac{P(x_i|Y=0)}{P(x_i|Y=1)} = \sum_i \ln \frac{\frac{1}{\sqrt{2\pi\sigma_{i0}^2}} \exp\left(\frac{-(x_i - \mu_{i0})^2}{2\sigma_{i0}^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_{i1}^2}} \exp\left(\frac{-(x_i - \mu_{i1})^2}{2\sigma_{i1}^2}\right)}$$

$$= \sum_i (\ln(\frac{1}{\sqrt{2\pi\sigma_{i0}^2}} \exp\left(\frac{-(x_i - \mu_{i0})^2}{2\sigma_{i0}^2}\right)) - \ln(\frac{1}{\sqrt{2\pi\sigma_{i1}^2}} \exp\left(\frac{-(x_i - \mu_{i1})^2}{2\sigma_{i1}^2}\right)))$$

$$= \sum_i (\ln \frac{1}{\sqrt{2\pi\sigma_{i0}^2}} - \frac{(x_i - \mu_{i0})^2}{2\sigma_{i0}^2} - \ln \frac{1}{\sqrt{2\pi\sigma_{i1}^2}} + \frac{(x_i - \mu_{i1})^2}{2\sigma_{i1}^2})$$

$$= \sum_i (\ln \frac{\sigma_{i1}}{\sigma_{i0}} + \frac{(x_i - \mu_{i1})^2}{2\sigma_{i1}^2} - \frac{(x_i - \mu_{i0})^2}{2\sigma_{i0}^2})$$

$$= \sum_i (\ln \frac{\sigma_{i1}}{\sigma_{i0}} + \frac{(\sigma_{i0}^2 - \sigma_{i1}^2){x_i}^2}{2\sigma_{i0}^2\sigma_{i1}^2} + \frac{(\mu_{i0}\sigma_{i1}^2 - \mu_{i1}\sigma_{i0}^2)x_i}{\sigma_{i0}^2\sigma_{i1}^2} + \frac{\sigma_{i0}^2{\mu_{i1}}^2 - \sigma_{i1}^2{\mu_{i0}}^2}{2\sigma_{i0}^2\sigma_{i1}^2})$$

其中有${x_i}^2$项，而 logistic 回归没有${x_i}^2$项，所以无法转换成 logistic 回归的形式。所以，一般高斯朴素贝叶斯分类器不是 logistic 回归所使用的形式。

- **Gaussian Bayes classifiers and logistic regression**

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

Answer:

根据上述非朴素的高斯贝叶斯分类器的假设，以及贝叶斯法则，有：

$$P(Y=1|X) = \frac{P(Y=1)P(X|Y=1)}{P(Y=1)P(X|Y=1) + P(Y=0)P(X|Y=0)}$$
$$= \frac{1}{1 + \frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}} = \frac{1}{1 + \exp\left(\ln \frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}\right)}$$
$$= \frac{1}{1 + \exp\left(\ln \frac{1-\pi}{\pi} + \ln \frac{P(X|Y=0)}{P(X|Y=1)}\right)}$$

其中$X = [x_1, x_2]^T$，且服从二元高斯分布：

$$P(x_1, x_2|Y=k) =$$

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

则：

$$\ln \frac{P(X|Y=0)}{P(X|Y=1)} = \ln \frac{P(x_1, x_2|Y=0)}{P(x_1, x_2|Y=1)}$$

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

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

$$=\left[\frac{\mu_{10}-\mu_{11}}{(1-\rho^2)\sigma_1^2}+\frac{\rho(\mu_{21}-\mu_{20})}{(1-\rho^2)\sigma_1\sigma_2}\right]x_1+\left[\frac{\mu_{20}-\mu_{21}}{(1-\rho^2)\sigma_2^2}+\frac{\rho(\mu_{11}-\mu_{10})}{(1-\rho^2)\sigma_1\sigma_2}\right]x_2 +\left[\frac{({\mu_{11}}^2-{\mu_{10}}^2)}{2(1-\rho^2)\sigma_1^2}+\frac{({\mu_{21}}^2-{\mu_{20}}^2)}{2(1-\rho^2)\sigma_2^2}+\frac{\rho(\mu_{10}\mu_{21}-\mu_{11}\mu_{21})}{(1-\rho^2)\sigma_1\sigma_2}\right]$$

那么等价于:

$$P(Y=1|X)=\frac{1}{1+\exp\left(w_0+\sum_{i=1}^{2}w_i x_i\right)}$$

其中

$$w_0=\ln\frac{1-\pi}{\pi}+\left[\frac{\left({\mu_{11}}^2-{\mu_{10}}^2\right)}{2(1-\rho^2)\sigma_1^2}+\frac{\left({\mu_{21}}^2-{\mu_{20}}^2\right)}{2(1-\rho^2)\sigma_2^2}+\frac{\rho\left(\mu_{10}\mu_{21}-\mu_{11}\mu_{21}\right)}{(1-\rho^2)\sigma_1\sigma_2}\right]$$

$$w_1=\frac{\mu_{10}-\mu_{11}}{(1-\rho^2)\sigma_1^2}+\frac{\rho\left(\mu_{21}-\mu_{20}\right)}{(1-\rho^2)\sigma_1\sigma_2}\qquad w_2=\frac{\mu_{20}-\mu_{21}}{(1-\rho^2)\sigma_2^2}+\frac{\rho\left(\mu_{11}-\mu_{10}\right)}{(1-\rho^2)\sigma_1\sigma_2}$$

满足 logistic 回归的一般形式，所以非朴素的高斯贝叶斯分类器也正是 logistic 回归所使用的形式。