Лекция 5. Распознавание образов, описываемых произвольными распределениями

Возможности использования гауссовской модели случайных векторов $\mathbf{x}$, описывающих образы в многомерном пространстве признаков, при всем ее удобстве, тем не менее, ограничены.

Во многих прикладных задачах для статистического описания образов следует использовать негауссовские модели данных. Решение о том, можно ли использовать гауссовскую модель или от нее следует отказаться, принимается на основе содержательного анализа ситуации и природы используемых признаков.

С другой стороны, постулирование любой другой статистической модели имеет смысл тогда, когда она позволяет реализовать конструктивное решающее правило.

Пусть для гипотез $\Omega = \{\omega_j,\ j = \overline{1,M}\}$ известны априорные вероятности гипотез и функции правдоподобия, каждая из которых описывается своим распределением $\omega_i : p(\omega_j)$, $p(x/\omega_i)$, $i = \overline{1,M}$ произвольного вида. Общая структура алгоритмов распознавания для случая проверки гипотезы о принадлежности любого образа одному из M классов объектов в этом случае ранее нами приведена. Напомним, что в исходном представлении она имеет вид

$$\omega_i:\ l_{ij}(x) = \frac{p(x/\omega_i)}{p(x/\omega_j)} \ge l_{ij} = \frac{p(\omega_j)}{p(\omega_i)},\quad j = \overline{1,M},\quad i \ne j$$

Конкретная реализация решающего правила и геометрия областей решений $\Gamma_i,\ i=\overline{1,M}$ определяются видом многомерных распределений вектора признаков каждого класса. Задание подобных распределений в аналитической форме при наличии статистической зависимости между признаками само по себе представляет весьма громоздкую задачу. Еще большие трудности возникают при попытках провести содержательный анализ получаемых решающих правил. Поэтому в данном случае либо используют непараметрические оценки распределений, о чем речь пойдет далее, либо вводят допущение о статистической независимости признаков, образующих вектор $\mathbf{x}$.

1. Распознавание образов в предположении статистической независимости признаков. Наивный байесовский классификатор

Использование предположения о статистической независимости признаков означает, что

$$p(x/\omega_i)=\prod_{k=1}^{n}p(x_k/\omega_i),\quad i=\overline{1,M}.$$

Следует понимать, что в случае, когда на самом деле это допущение не выполняется, использование синтезированного алгоритма может привести к потерям с точки зрения достигаемых характеристик качества распознавания. Поэтому такое допущение можно принимать в случае, когда признаки либо реально независимы, либо «почти независимы» о чем можно судить, проводя содержательный (в том числе и математический) анализ природы этих признаков.

Синтезируемый на основе такого представления алгоритм распознавания в литературе называется **наивным байесовским классификатором**. При его использовании в задаче синтеза достигается **главное преимущество**: возможность записи в аналитическом виде произвольной одномерной плотности распределения вероятностей $p(x_k/\omega_i)$, описывающей поведение конкретного признака, используемого в качестве компонента вектора $\mathbf{x}$.

Рассмотрим случай распознавания двух классов образов. В этом случае решающее правило с учетом проведения логарифма отношения правдоподобия может быть представлено в виде

$$\ln l(x) = g''(x) = \sum_{k=1}^{n} \ln \frac{p(x_k/\omega_1)}{p(x_k/\omega_2)} = \sum_{k=1}^{n} g_k''(x_k) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0', \qquad l_0' = \ln l_0 .$$

Используя полученное соотношение, перейдем к анализу решающего правила. Для этого обычно проводят аппроксимацию распределения ЛОП в разделяющей функции гауссовскими распределениями $\mathbf{g}''(\mathbf{x})$: $p(g''/\omega_1) \approx N(g'', m_{g1}, D_{g1})$, $p(g''/\omega_2) \approx N(g'', m_{g2}, D_{g2})$.

Очевидно, что теперь для использования подобного приема больше оснований, чем, например, в случае, когда рассматривается задача распознавания ГСВ с различными матрицами ковариациями. Это связано с тем, что величина $\mathbf{g}''(\mathbf{x})$ здесь есть сумма величин ЛОП, определяемых по каждому признаку в отдельности.

Такая сумма, в соответствие с центральной предельной теоремой при достаточно большом n и отсутствии доминирования распределения одного из используемых признаков над другими, сходится к гауссовскому распределению.

При использовании гауссовской аппроксимации указанных плотностей необходимо рассчитать их математические ожидания и дисперсии

$$m_{gi} = M[\mathbf{g}''(\mathbf{x}) / \omega_i] = \sum_{k=1}^{n} m_{gk,i}\ ,\ D_{gi} = M[(\mathbf{g}''(\mathbf{x}) - m_{gi})^2 / \omega_i] = \sum_{k=1}^{n} D_{gk,i}\ ,\ i = 1{,}2$$

$$m_{gk,i} = M[\mathbf{g}_k''(\mathbf{x}_k) / \omega_i] = \int g_k''(x_k) p(x_k / \omega_i) dx_k\,,$$

$$D_{gk,i} = M[(\mathbf{g}_k''(\mathbf{x}_k) - m_{gk,i})^2 / \omega_i] = \int (g_k''(x_k) - m_{gk,i})^2 p(x_k / \omega_i) dx_k\,.$$

При известной плотности распределения каждого признака такие расчеты выполняются относительно несложно в аналитическим виде или путем численного интегрирования. Рассмотрим в качестве примера случай, когда признаки распределены по показательному закону с одним и тем же параметром в пределах каждого класса

$$p(x_k / \omega_1) = \frac{1}{\lambda_1} \exp\left(-\frac{x_k}{\lambda_1}\right), \quad k = \overline{1,n}\,,\ p(x_k / \omega_2) = \frac{1}{\lambda_2} \exp\left(-\frac{x_k}{\lambda_2}\right), \quad k = \overline{1,n}\,.$$

Математические ожидания и дисперсии этих распределений равны

$$m_{xk,i} = M[\mathbf{x}_k / \omega_i] = \lambda_i, \quad D_{xk,i} = M[(\mathbf{x}_k - m_{xk,i})^2 / \omega_i] = \lambda_i^2, \quad k = \overline{1,n}, \ i = 1{,}2\,,$$

Соответственно, для параметров распределений слагаемых разделяющей функции получим

$$g_k''(x_k) = \ln\left(\frac{1}{\lambda_1} \exp\left(-\frac{x_k}{\lambda_1}\right)\right) \bigg/ \left(\frac{1}{\lambda_2} \exp\left(-\frac{x_k}{\lambda_2}\right)\right) = \ln\frac{\lambda_2}{\lambda_1} + x_k\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right),$$

$$m_{gk,i} = \ln\frac{\lambda_2}{\lambda_1} + \lambda_i\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right), D_{gk,i} = \lambda_i^2\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^2, \quad k = \overline{1,n}\,,$$

$$m_{gi} = \sum_{k=1}^{n} m_{gk,i} = n\left[\ln\frac{\lambda_2}{\lambda_1} + \lambda_i\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)\right], i = 1,2$$

$$D_{gi} = \sum_{k=1}^{n} D_{gk,i} = n\left[\lambda_i^2\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^2\right], \quad i = 1,2$$

На их основе окончательно получим выражения для вероятностей ошибок первого и второго рода в виде

$$\alpha = \int_{-\infty}^{l_0'} N(g'', m_{g1}, D_{g1})dg'' = \Phi\left(\frac{l_0' - m_{g1}}{\sqrt{D_{g1}}}\right), \ \beta = \int_{l_0'}^{\infty} N(g'', m_{g2}, D_{g2})dg'' = 1 - \Phi\left(\frac{l_0' - m_{g2}}{\sqrt{D_{g2}}}\right).$$

Алгоритм распознавания **для случая многих классов** с учетом результатов, полученных для двух классов, может быть проанализирован по аналогичной ранее рассмотренным схеме.

2. Распознавание образов в случае статистически независимых дискретных признаков

Одним из важных случаев, возникающих на практике, является случай, когда исходные признаки являются дискретными, т.е. принимают дискретный ряд значений.

Пусть, для определенности, вектор признаков $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ состоит из компонент, каждая из которых принимает дискретный ряд значений $\mathbf{x}_i \in V = \{v_1, \ldots, v_m\}, k = \overline{1, n}$. В простейшем случае множество значений состоит из двух элементов $\mathbf{x}_k \in V = \{0,1\}$. Такие признаки будем назвать бинарными.

Пусть также для каждого класса образов заданы априорные вероятности гипотез $\omega_i : p(\omega_j)$ и функции правдоподобия, каждая из которых может быть представлена в виде $p(x/\omega_i),\ i=\overline{1,M}$

$$p(x/\omega_i) = \prod_{k=1}^{n} p(x_k/\omega_i),\quad i=\overline{1,M}\,,$$

$$p(x_k/\omega_i) = \sum_{t=1}^{m} p_k(v_t/\omega_i)\delta(x_k - v_t),\quad \sum_{t=1}^{m} p_k(v_t/\omega_i) = 1,\quad k=\overline{1,n}\,,\ i=\overline{1,M}\,,$$

где $p_k(v_t/\omega_i) = p_{k,t,i} = p_{k,t}(\omega_i)$ – различные обозначения вероятности получения значения v_t для признака $\mathbf{x}_k$ при условии, что образ принадлежит классу ω_i. Здесь статистические свойства дискретной случайно величины представлены с помощью плотности распределения общего вида с использованием дельта-функций Дирака.

Другая возможная форма записи использует дискретные дельта-функции Кронекера

$$p(x_k/\omega_i) = \sum_{t=1}^{m} p_k(v_t/\omega_i)\tilde{\delta}(x_k - v_t)\,,\ \tilde{\delta}(x_k - v_t) = \begin{cases} 1, & x_k = v_t, \\ 0, & x_k \neq v_t. \end{cases}$$

Рассмотрим задачу распознавания классов с использованием независимых бинарных признаков, принимающих значения с вероятностью 0 и 1. Такая модель подразумевает, что каждый из признаков несёт ответ типа «да» или «нет». При описании каждого класса и каждого признака в нем используются только две вероятности

$$p_{k,i} = p(x_k = 1/\omega_i)\,,\ \ 1 - p_{k,i} = p(x_k = 0/\omega_i)\,,\ \ k=\overline{1,n}\,.$$

Полное вероятностное описание каждого класса имеет форму

$$p(x/\omega_i) = \prod_{k=1}^{n} p_{k,i}^{x_k} (1 - p_{k,i})^{1-x_k}, \quad i = \overline{1,M}.$$

В качестве примера использования такой модели для описания реальной ситуации может служить модель распознавания бинарных изображений в условиях воздействия помех, которые приводят к инвертированию отдельных элементов (пикселей) изображения, изменяя их значения на противоположные с одинаковой вероятностью $p_I > 0$.

Рис.1. Исходные изображения и их искаженные шумом версии

На рисунке 1 представлены изображения двух букв, являющихся матрицами размера 5×7, состоящими из нулей и единиц, а также их искаженные версии, полученные в результате воздействия помехи, независимо инвертирующей элементы исходных образов с вероятностью $p_I = 0.1$.

Выполним синтез и анализ алгоритма распознавания в случае двух классов, имея в виду, что переход к случаю многих классов здесь может быть произведен аналогичным образом, как и во всех ранее рассмотренных задачах.

Введем более удобные обозначения для вероятностей значений бинарных признаков двух классов. Пусть

$$p_k = p(x_k = 1/\omega_1),\; 1 - p_k = p(x_k = 0/\omega_1),\; k = \overline{1,n},$$

$$q_k = p(x_k = 1/\omega_2),\; 1 - q_k = p(x_k = 0/\omega_2),\; k = \overline{1,n}.$$

Выражения для функций правдоподобия классов и ЛОП можно записать в виде

$$p(x/\omega_1) = \prod_{k=1}^{n} p_k^{x_k}(1-p_k)^{1-x_k},\; p(x/\omega_2) = \prod_{k=1}^{n} q_k^{x_k}(1-q_k)^{1-x_k},$$

$$g''(x) = \sum_{k=1}^{n} \ln\frac{p(x_k/\omega_1)}{p(x_k/\omega_2)} = \sum_{k=1}^{n}\left(x_k \ln\frac{p_k}{q_k} + (1-x_k)\ln\frac{1-p_k}{1-q_k}\right) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0'. \tag{1}$$

Анализ полученного выражения показывает, что ЛОП и разделяющая функция $g'(x) = g''(x) - l_0'$ является линейной функцией от компонент x, образуя линейную границу между класс в виде

$$g'(x) - l_0' = \sum_{k=1}^{n} \alpha_k x_k + \alpha_0 = 0,$$

$$\alpha_k = \ln\frac{p_k(1-q_k)}{q_k(1-p_k)},\; \alpha_0 = \sum_{k=1}^{n}\ln\frac{1-p_k}{1-q_k} - l_0'.$$

Для расчета вероятностей ошибок распознавания в данной задаче могут использоваться приближенные соотношения для законов распределения $g''(x)$ или прямое имитационное моделирование алгоритма.

Рассмотрим частный случай, часто встречающийся в прикладных задачах распознавания по бинарным признакам. Он связан с тем, что для всех признаков каждого из двух классов вероятности единиц и нулей одинаковы $p_k = p \neq 0$, $q_k = q \neq 0$, $k = \overline{1,n}$.

Такая ситуация означает, что проводится «опрос» n независимых и равноценных признаков или, что, эквивалентно, n раз проводятся независимые наблюдения одного и того же признака. Тогда формула (1) с предыдущей с страницы преобразуется так

$$g''(x) = L_x \ln\frac{p}{q} + P_x \ln\frac{1-p}{1-q} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0', \quad L_x = \sum_{k=1}^{n} x_x, \quad P_x = \sum_{k=1}^{n}(1-x_x) = n - L_x,$$

где L_x, P_x – количество единиц (ответов «да») и количество нулей (ответов «нет»), полученные в ходе наблюдения. Пусть для определенности причем $p > q$, тогда можно преобразовать следующим образом:

$$\ln\frac{p(1-q)}{q(1-p)} \sum_{k=1}^{n} x_k \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0' - n\ln\frac{1-p}{1-q}$$

что эквивалентно

$$L_x = \sum_{k=1}^{n} x_k \underset{\omega_2}{\overset{\omega_1}{\gtrless}} L_0 = round\left[\left(l_0' - n\ln\frac{1-p}{1-q}\right) \Big/ \ln\frac{p(1-q)}{q(1-p)}\right].$$

Отметим, что при $p > q$ величина $\ln\frac{p(1-q)}{q(1-p)}$ больше нуля и знак неравенства не меняется.

Полученное решающее правило эквивалентно сравнению количества полученных единиц с порогом, вычисляемым на основе априорных вероятностей гипотез и значения вероятностей единиц и нулей обоих классов.

Такое представление алгоритма позволяет, используя схему Бернулли для описания n независимых испытаний, в данном случае записать точные выражения для вероятностей ошибок распознавания в виде

$$\alpha = \sum_{t=0}^{L_0-1} C_n^t p^t (1-p)^{n-t}, \qquad \beta = 1 - \sum_{t=0}^{L_0} C_n^t q^t (1-q)^{n-t}.$$

Пример: использование алгоритма при обнаружении импульсной последовательности сигналов.

Пример решения задачи распознавания изображений по бинарным признакам. Для каждого класса здесь вероятностные описания признаков с учетом заданной вероятности искажения помехой $0 < p_I < 1$ можно записать в виде

$$p_{k,i} = p(x_k = 1/\omega_i) = s_k^{(i)}(1-p_I) + (1-s_k^{(i)})p_I,$$

$$1 - p_{k,i} = p(x_k = 0/\omega_i) = s_k^{(i)} p_I + (1-s_k^{(i)})(1-p_I), \; k = \overline{1,n}, i = \overline{1,M}.$$

Здесь $s_k^{(i)}, k=\overline{1,n}$ – элементы исходного, неискаженного бинарного изображения ω_i, развернутого в вектор - столбец, принимающие значения единица или ноль в зависимости от того, закрашен ли ($s_k^{(i)}=1$) соответствующий элемент изображения или нет ($s_k^{(i)}=0$). Т.е. в данной задаче вероятностные характеристики значений двоичных признаков для каждого класса различны.

Алгоритм распознавания на основе сравнения ЛОП с порогом в случае двух классов (или попарного сравнения классов в случае $M>2$) в данном случае имеет вид

$$p(x/\omega_1)=\prod_{k=1}^{n}[s_k^{(1)}(1-p_I)+(1-s_k^{(1)})p_I]^{x_k}[s_k^{(1)}p_I+(1-s_k^{(1)})(1-p_I)]^{1-x_k},$$

$$p(x/\omega_2)=\prod_{k=1}^{n}[s_k^{(2)}(1-p_I)+(1-s_k^{(2)})p_I]^{x_k}[s_k^{(2)}p_I+(1-s_k^{(2)})(1-p_I)]^{1-x_k},$$

$$g''(x)=\sum_{k=1}^{n}\left(x_k\ln\frac{s_k^{(1)}(1-p_I)+(1-s_k^{(1)})p_I}{s_k^{(2)}(1-p_I)+(1-s_k^{(2)})p_I}+(1-x_k)\ln\frac{s_k^{(1)}p_I+(1-s_k^{(1)})(1-p_I)}{s_k^{(2)}p_I+(1-s_k^{(2)})(1-p_I)}\right)\underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0'.$$

В случае, если $p_I=0$, как и в случае, если $p_I=1$, задача становится вырожденной, так как

$$p(x/\omega_i)=\prod_{k=1}^{n}\left(s_k^{(i)}\right)^{x_k}\left(1-s_k^{(i)}\right)^{1-x_k}=\begin{cases}1, & x=s^{(i)}\\ 0, & x\neq s^{(i)}\end{cases},\quad i=\overline{1,2},$$

и весовые коэффициенты в выражение для разделяющей функции могут обратиться в бесконечность.

Так же в случае, если $p_I=0.5$,

$$p(x/\omega_i) = \prod_{k=1}^{n} 0.5^{x_k} 0.5^{1-x_k} = 0.5^n, \; i = \overline{1,2}$$

и, независимо от значения x, всегда $g''(x) \equiv 0$. В этой ситуации решение принимается в пользу класса, имеющего наибольшую априорную вероятность.

Для проведении анализа решающего правила учтем, что в некоторых элементах (пикселях) исходные бинарные изображения совпадают, т.е. одновременно $s_k^{(1)} = 1$, $s_k^{(2)} = 1$ или $s_k^{(1)} = 0$, $s_k^{(2)} = 0$. Эти точки из вычисления ЛОП исключаются, что позволяет записать

$$g''(x) = L_{x,10} \ln \frac{1-p_I}{p_I} + L_{x,01} \ln \frac{p_I}{1-p_I} + P_{x,01} \ln \frac{1-p_I}{p_I} + P_{x,10} \ln \frac{p_I}{1-p_I},$$

где, соответственно, $L_{x,10}, L_{x,01}, P_{x,01}, P_{x,10}$: количество полученных единиц в тех элементах, где $s_k^{(1)} = 1$ и $s_k^{(2)} = 0$; количество полученных единиц в тех элементах, где $s_k^{(1)} = 0$ и $s_k^{(2)} = 1$; количество полученных нулей в тех элементах, где $s_k^{(1)} = 0$ и $s_k^{(2)} = 1$; количество полученных нулей в тех элементах, где $s_k^{(1)} = 1$ и $s_k^{(2)} = 0$.

Определив общее количество несовпадающих элементов двух изображений, как $n_s = \sum_{k=1}^{n} \left| s_k^{(1)} - s_k^{(2)} \right|$, еще раз преобразуем последнее выражение приведя его к виду

$$g''(x) = (L_{x,10} + P_{x,01}) \ln \frac{1-p_I}{p_I} + (n_s - L_{x,10} - P_{x,01}) \ln \frac{p_I}{1-p_I} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0',$$

Теперь, если $0 < p_I < 0.5$ и, соответственно, $\ln(1-p_I)/p_I > 0$ получим эквивалентное неравенство

$$g''(x) = \left(L_{x,10} + P_{x,01}\right) \mathop{\gtrless}_{\omega_2}^{\omega_1} L_0 = \frac{l_0'}{2\ln(1-p_I) - 2\ln p_I} + \frac{n_s}{2}.$$

Отсюда вероятности ошибок для определяются как

$$\alpha = \sum_{t \ge L_0}^{n_s} C_{n_s}^t (1-p_I)^t p_I^{n_s - t}, \quad \beta = 1 - \sum_{t \ge L_0}^{n_s} C_{n_s}^t p_I^t (1-p_I)^{n_s - t}.$$

Если же $0.5 < p_I < 1$, то $\ln(1-p_I)/p_I < 0$ и эквивалентное неравенство имеет вид

$$g''(x) = \left(L_{x,10} + P_{x,01}\right) \mathop{\lessgtr}_{\omega_2}^{\omega_1} L_0 = \frac{l_0'}{2\ln(1-p_I) - 2\ln p_I} + \frac{n_s}{2},$$

а его вероятностные характеристики окончательно записываются в виде

$$\alpha = 1 - \sum_{t \ge L_0}^{n_s} C_{n_s}^t (1-p_I)^t p_I^{n_s - t}, \quad \beta = \sum_{t \ge L_0}^{n_s} C_{n_s}^t p_I^t (1-p_I)^{n_s - t}.$$

На рис.2а, б представлены зависимости для теоретических вероятностей ошибок и их экспериментальных оценок, полученных в ходе статистического моделирования.

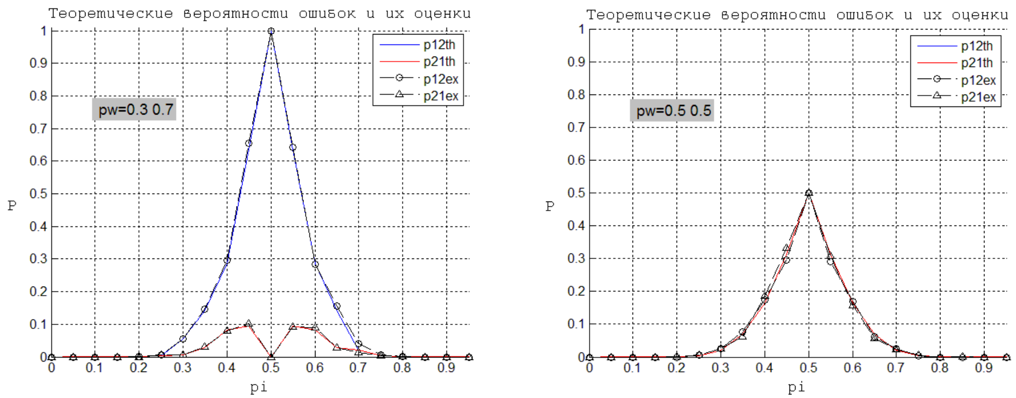

Рис.2. Зависимости для вероятностей ошибок первого и второго рода о вероятности искажения элементов p_I

Анализ полученных зависимостей показывает, прежде всего, хорошее совпадение теоретических и экспериментальных результатов. Видно, что при увеличении величины p_I до уровня 0.5 происходит монотонный рост вероятностей ошибок. В точке 0.5, когда ЛОП равен нулю, решение принимается в пользу того класса, который имеет большую априорную вероятность. При дальнейшем увеличении p_I происходит симметричное снижение вероятностей ошибок за счет возникновения «переинверсии. При этом качество распознавания повышается, что эквивалентно использованию в качестве эталонов инвертированных версий изображений.