ТУБ 5

5.1 Понятие вероятности. Теорема Бейеса

Поскольку большинство происходящих событий носят случайный характер, представляют интерес методы их классификации. Случайное событие – это такое, для которого невозможно предсказать его точный результат в каждом конкретном случае. Однако при большом числе реализаций эти события можно характеризовать средними результатами, стабильными и достоверными. В основе представлений случайных событий лежит понятие вероятности.

Рассмотрим опыт с несколькими возможными исходами, каждый конкретный результат которого в точности не предсказуем. Множество всех исходов назовем Ω – пространство элементарных событий. Событие A есть подмножество множества Ω, обладающее некоторыми свойствами, приводимыми ниже.

По определению вероятностью *P(A)* события A называют величину, удовлетворяющую трем аксиомам:

1) $0 \le P(A) \le 1$;
2) $P(\Omega) = 1$;
3) если $\{A_i\}$ $-$ счетное множество событий,
такое, что $\mathrm{A}_i \cap \mathrm{A}_j = \Phi \; \forall i, j$, то $P(\bigcup_i A_i) = \sum_i P(A_i)$.

Здесь через Φ обозначено пустое множество. К этому можно добавить, что $P(\bar{A}) = 1 - P(A)$, где $\bar{A}$ – событие, дополняющее A до Ω.

Если могут произойти два события – A и B, то говорят о трех различных вероятностях:

событие A происходит с вероятностью $P(A)$,

событие B происходит с вероятностью $P(B)$,

события A и B происходят одновременно с вероятностью $P(A, B)$, ее называют вероятностью совместного события, или совместной вероятностью.

Пусть теперь A_i (i=1, 2, … n) и B – случайные события. Условную вероятность того, что произойдет событие A_i, при условии, что произошло событие B, записывают так: $P(A_i/B)$. Это – апостериорная вероятность. Ее плотность обозначают $p(A_i/B)$.

Величину условной вероятности $P(A_i/B)$ можно вычислить по теореме

Бейеса: $P\left(\frac{A_i}{B}\right) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)p\left(\frac{B}{A_i}\right)}{\sum_i P(A_j)p\left(\frac{B}{A_j}\right)}$, поскольку $P(B) = \sum_i p(B/A_j)P(A_j)$.

Действительно, $P(B) = P(A_1B) + P(A_2B) + \ldots + P(A_nB)$, а при этом

$P(A_iB) = P(A_i)P\left(\frac{B}{A_i}\right)$. Здесь $P(A_i)$ – априорная вероятность события A_i.

Примером апостериорной плотности вероятности является случай одномерного гауссового распределения, выражаемого формулой

$$p(x/j) = \frac{1}{\sigma_j\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2].$$

Эта плотность распределения является функцией двух параметров: μ_j– математическое ожидание и σ_j – среднеквадратичное отклонение. Параметры могут быть вычислены по N опытам, в каждом из которых измеряются величина x_k $(k = 1, 2, \ldots N)$: $\hat{\mu}_j = \frac{1}{N}\sum_{k=1}^{N} x_k$; $\hat{\sigma}_j^2 = \frac{1}{N}\sum_{k=1}^{N}(x_k - \hat{\mu}_j)^2$.

5.2 Решающее правило

С помощью понятия об апостериорных вероятностях можно подойти к разработке метода автоматической классификации.

Пусть Δ – сепарабельное пространство признаков, а $\vec{X}$ – вектор, представляющий κ-ый класс, сепарабельное пространство по определению может быть разделено на классы. Априорная вероятность того, что данная реализация относится к классу с номером κ, есть $P(\vec{X}_k)$. Она считается заданной самой постановкой задачи.

Задача заключается в том, чтобы отнести неизвестный предъявляемый объект $\vec{X}$ к одному из известных классов C_k с минимальной ошибкой. Для этого выполняют n измерений в соответствии с признаками, выбранными надлежащим образом. В результате получают вектор измерений $\vec{X}_m$, для которого можно найти условную вероятность или ее плотность: $p(\vec{X}_m/C_k)$.

Решение об отнесении неизвестного объекта к классу с номером κ можно считать оправданным, если для любого j выполняется условие:
$p(\vec{X}_m/C_k) \geq p(C_j/\vec{X}_m)\ \forall j$. Эти вероятности могут быть вычислены согласно теореме Бейеса по тем условным вероятностям $p(\vec{X}_m/C_k)$, которые получаются непосредственно в процессе измерений:

$$P(C_k/\vec{X}_m) = \frac{P(C_k)p(\vec{X}_m/C_k)}{p(X_m)}, \quad P(C_j/\vec{X}_m) = \frac{P(C_j)p(\vec{X}_m/C_j)}{p(X_m)},$$

Откуда следует решающее правило:

$$P(C_k)p(\vec{X}_m/C_k) \geq P(C_j)p(\vec{X}_m/C_j).$$

Процесс разделения на классы, как правило, связан с некоторым риском. Правильной или неверной классификации можно придать определенную «цену». Введем решающее правило d, в соответствии с которым каждый предъявляемый объект $\vec{X}_m$ по результатам измерений вызывает действие d_i. Оно определяется отнесением $\vec{X}_m$ к классу C_i. Будем полагать, что кроме заранее установленных классов, для сомнительных случаев отведен дополнительный класс.

Решающее правило можно записать в виде $\vec{X}_m \xrightarrow{d(\vec{X}_m)} \{d_i\},\ i = 1,2, \dots l$. Решение d_i означает, что объект $\vec{X}_m$ отнесен к классу C_i. Но это не означает, что данное решение бесспорно, т.е. $\vec{X}_m$ может как принадлежать классу C_i, так и не принадлежать ему.

Приведем в соответствие каждому решению некоторую цену или степень риска. Пусть $c(d_i/C_j)$ – условная цена принятия решения d_i, если известно, что объект принадлежит классу C_j. Среднее значение цены, связанной с принятием решения d_i в случае, когда предъявлен объект $\vec{X}_m$, определяет средний риск принятия данного решения: $R(d_i/\vec{X}_m) = \sum_{j=1}^{k} c(d_i/C_j)P(C_j/\vec{X}_m)$, где правая часть равенства получена по теореме Бейеса.

Средняя цена для решающего правила d, если известно, что предъявлен объект $\vec{X}_m$, есть $R(d/\vec{X}_m) = \sum_{j=1}^{k} c(d/C_j)P(C_j/\vec{X}_m)$.

Общая средняя цена, связанная с решающим правилом d для всех реализаций $\vec{X}_m$, будет $R(d) = \int_{\vec{X}_m} R(d/\vec{X}_m)p(\vec{X}_m)d\vec{X}_m$.

Таким образом, классификация сводится к отысканию такого оптимального решающего правила d^*, которое минимизировало бы общее среднее значение цены: $R^* = R(d^*) = \underset{d}{Min} R(d)$.

Пусть $c_{ij} = c(d_i/C_j)$ – цена, связанная с решением отнести $\vec{X}_m$ к классу C_i, если известно, что искомый класс есть C_j. Функцию цены в этом случае представим в следующем виде:

$$c_{1j} = 1 - \delta_{ij} \begin{cases} = 0 \text{ при } i = j \\ = 1 \text{ при } i \neq j \end{cases}, \text{ так как } \delta_{ij} = \begin{cases} 1 \text{ при } i = j \\ 0 \text{ при } i \neq j \end{cases}.$$

Для данного $\vec{X}_m$ средняя цена отнесения его к классу C_i будет: $R_i(\vec{X}_m) = \sum_{j=1}^{k} c_{ij}P(C_j/\vec{X}_m)$, тогда $R_i(\vec{X}_m) = 1 - P(C_i/\vec{X}_m)$, и решающее правило преобразуется к виду $d^*(\vec{X}_m) = d_i \Leftrightarrow P(C_i/\vec{X}_m) = \underset{j=1,k}{Max} P(C_j/\vec{X}_m)$, поскольку стремятся минимизировать цену $R_i(\vec{X}_m)$. При этом вероятность ошибочного решения

$$R^* = \int_{\vec{X}_m} R^*(\vec{X}_m)p(\vec{X}_m)d\vec{X}_m \quad \text{или} \quad R^*(\vec{X}_m) = 1 - \underset{j=1,k}{Max} P(C_j/\vec{X}_m).$$

5.3 Разделение объектов на два класса при вероятностном подходе

Рассмотрим случай, когда весь набор возможных решений сводится к двум, т.е. предъявленный объект может быть отнесен к одному из двух имеющихся классов. При этом полагаем, что установлены средние цены $R_1(\vec{X}_m)$ и $R_2(\vec{X}_m)$: $R_1(\vec{X}) = c_{11}P(C_1/\vec{X}_m) + c_{12}P(C_2/\vec{X}_m)$,

$$R_2(\vec{X}) = c_{21}P(C_1/\vec{X}_m) + c_{22}P(C_2/\vec{X}_m).$$

Требуется найти такое решение d_i, которое минимизировало бы среднюю цену:

$d(\vec{X}_m) = C_1 \Rightarrow R_1(\vec{X}) \le R_2(\vec{X})$ или более подробно
$c_{11}P(C_1/\vec{X}_m) + c_{12}P(C_2/\vec{X}_m) \le c_{21}P(C_1/\vec{X}_m) + c_{22}P(C_2/\vec{X}_m)$.

Положим $c_{11} < c_{12}$ и $c_{22} < c_{21}$, что соответствует случаю, когда риск принятия верного решения меньше риска допустить ошибку. Тогда из предыдущего условия $(c_{11} - c_{21})P(C_1/\vec{X}_m) \le (c_{22} - c_{12})P(C_2/\vec{X}_m)$ имеем

$(c_{21} - c_{11})P(C_1/\vec{X}_m) \ge (c_{12} - c_{22})P(C_2/\vec{X}_m)$. Откуда $\frac{P(C_1/\vec{X}_m)}{P(C_2/\vec{X}_m)} \ge \frac{c_{12}-c_{22}}{c_{21}-c_{11}}$.

Применяя формулу Бейеса, получим решающее правило в виде неравенства

$$\frac{p(\vec{X}_m/C_1)}{p(\vec{X}_m/C_2)} \ge \frac{(c_{12} - c_{22})P(C_2)}{(c_{21} - c_{11})P(C_1)}.$$

Дробь, стоящую в левой части неравенства, называют отношением правдоподобия. Решение принимают тогда, когда эта величина достигает максимума. Правая часть неравенства, содержащая только известные члены, представляет собой величину порога, не зависящего от наблюдаемой величины.

Если представить цену в виде $c_{ij} = 1 - \delta_{ij}$, то условия $d(\vec{X}_m) = d_i \Leftrightarrow P(C_i/\vec{X}_m) = \underset{j=1,k}{Max} P(C_j/\vec{X}_m)$ преобразуется в $d(\vec{X}_m) = d_1 \Leftrightarrow P(C_1/\vec{X}_m) \ge P(C_2/\vec{X}_m)$ или после применения формулы Бейеса

$$\frac{p(\vec{X}_m/C_1)}{p(\vec{X}_m/C_2)} \ge \frac{P(C_2)}{P(C_1)}.$$

Рассмотрим вероятности ошибок, которые могут возникать при такой процедуре. На рисунке 1 показаны две кривые, представляющие зависимость $p(\vec{X}_m/C_i)P(C_i)$ при i=1,2. Очевидно, что на прямой AB, неравенство Бейеса выполняется, и можно заключить, что $\vec{X}_m$ принадлежит классу C_1.

Рассмотрим линию раздела, обозначенную Δ. Любая точка, для которой $\vec{X}_m < X$, считается принадлежащей классу C_1, в то время как все точки, для которых $\vec{X}_m > X$, относятся к классу C_2. Однако вероятность того, что в первом случае точка может принадлежать классу C_2, отлична от нуля (область 1), так же как и то, что во втором случае точка X принадлежит классу C_1 (область 2). Для класса C_1 зона 1 является зоной ложной тревоги, а зона 2 является зоной пропуска обнаружения. Они определяются соответственно выражениями:

$P_{\text{л.т.}} = \int_{-\infty}^{x} P(C_2)p(\vec{X}_m/C_2)\, d\vec{X}_m \qquad P_{\text{п.о.}} = \int_{x}^{\infty} P(C_1)p(\vec{X}_m/C_1)\, d\vec{X}_m.$

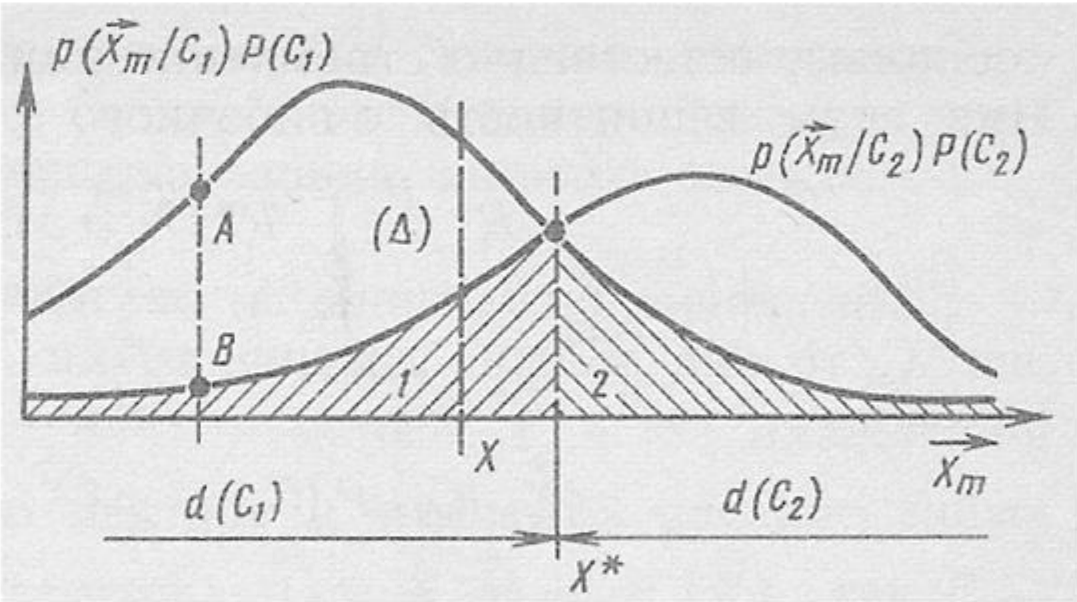

Рисунок 1 – Вероятности ошибок при разделении на два класса

Суммарная ошибка классификации представляется суммой этих двух вероятностей. Если перемещать линию Δ, разделяющую два решения, вдоль оси X, то она должна достичь точки X^*, в которой имеет место равенство $P(C_1)p(\vec{X}_m/C_1) = P(C_2)p(\vec{X}_m/C_2)$, показывающее, что при бинарных ценах правило максимума правдоподобия обеспечивает оптимальную классификация по отношению к возможности ошибочного решения.