Лекция 15

МЕТОДЫ РАЗДЕЛЕНИЯ В ПРОСТРАНСТВЕ ИНФОРМАТИВНЫХ ПАРАМЕТРОВ

Дискриминантные и разделяющие функции

Если образы — векторы обучающей выборки, образуют компактные области в пространстве информативных параметров, можно воспользоваться метрическими методами постановки диагноза.

Однако возможна такая ситуация, когда векторы, соответствующие различным состояниям, в пространстве информативных параметров занимают довольно обширные зоны, что делает применение метрических методов невозможным. На рис. Л15.1 показано распределение образов объектов, находящихся в одном из двух возможных состояний. Видно, что они распределены в двух областях информативного пространства.

Задача диагностирования в этом случае сводится к нахождению поверхностей (линий), разделяющих образы разных состояний, и определению, в какую из областей попадает вектор диагностируемого объекта.

Рассмотрим n-мерное пространство информативных параметров. Если диагностируемый объект может находиться в одном из $N+1$ состояниях $D_0, D_1, ..., D_N$, то пространство содержит точки, принадлежащие N+1 различным диагнозам. *Областью диагноза* D_k $\left(k = 0, 1, ..., N\right)$ называется множество точек пространства, соответствующих состоянию объекта D_k.

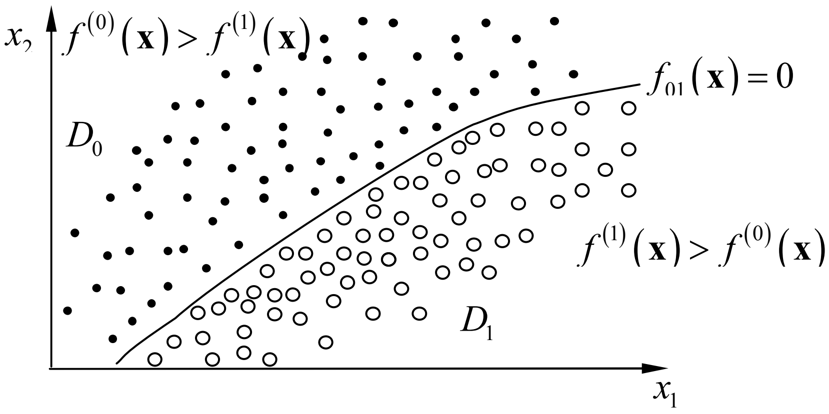

Рис. Л15.1. Распределение в информативном пространстве образов объектов, которые могут находиться в одном из двух состояний D_0 и D_1

В дальнейшем будем полагать, что информативные параметры нормированы и размерность их равна единице. С целью решения задачи диагностирования в пространстве параметров задается множество функций $\left\{f^{(k)}(\mathbf{x})\right\}$ $(k = 0,1,...,N)$ для каждой области диагноза, таких, что если вектор $\mathbf{x} \in D_m$, то

$$f^{(p)}(\mathbf{x}) > f^{(k)}(\mathbf{x}), \quad k = 0,1,...,N; \; (p \neq k). \tag{Л15.1}$$

Для вектора $\mathbf{x}$, установленного в результате обследования объекта, $f^{(p)}(\mathbf{x})$ принимает наибольшее значение по сравнению с другими функциями $f^{(k)}(\mathbf{x})$ $(p \neq k)$, если объект находится в состоянии D_p. Определенные таким образом функции называются *дискриминантными*.

Если области диагнозов D_i и D_m имеют общую границу, то можно ввести *разделяющую функцию* этих состояний

$$f_{ip}(\mathbf{x}) = f^{(i)}(\mathbf{x}) - f^{(p)}(\mathbf{x}), \tag{Л15.2}$$

тогда уравнение разделяющей поверхности областей диагнозов имеет вид:

$$f_{ip}(\mathbf{x}) = 0. \tag{Л15.3}$$

На рис. Л15.2 показано разграничение двухмерного информативного пространства разделяющей функцией $f_{01}(\mathbf{x}) = 0$ на две области D_0 и D_1, в каждой из которых определены дискриминантные функции $f^{(0)}(\mathbf{x})$ и $f^{(1)}(\mathbf{x})$.

Правило принятия решения о состоянии диагностируемого объекта формулируется следующим образом: если для вектора $\mathbf{x}$ выполняется соотношение (Л15.1) или эквивалентное ему неравенство

$$f_{ip}(\mathbf{x}) > 0, \;\; p = 0,1,..., N; \;\; (i \neq p) \tag{Л15.4}$$

то ставится диагноз $D_i \; (\mathbf{x} \in D_i)$.

Для повышения надежности диагностирования вводится зона неопределенности. С этой целью задают малую положительную величину ε такую, что

при $f_{ip}(\mathbf{x}) > \varepsilon, \;\; \mathbf{x} \in D_i, \;\; p = 0,1,..., N; \;\; (i \neq p)$;

при $\left|f_{ip}(\mathbf{x})\right| < \varepsilon$ отказываются от распознавания.

Выбор дискриминантных или разделяющих функций составляет основу *методов разделения*. Наиболее общим является представление дискриминантных и разделяющих функций в виде рядов

$$f^{(k)}\left(\mathbf{x}\right)=\sum_{j=0}^{m_1}c_j^{(k)}\varphi_j\left(\mathbf{x}\right); \tag{Л15.5}$$

$$f_{ik}(\mathbf{x})=\sum_{j=0}^{m_2}c_j^{(ik)}\varphi_j(\mathbf{x}), \tag{Л15.6}$$

где $c_j^{(k)}$, $c_j^{(ik)}$ — постоянные коэффициенты. Выбрав соответствующим образом систему функций $\left\{\varphi_j(\mathbf{x})\right\}$ и полагая число слагаемых в суммах m_1 или m_2 достаточно большими, в пределе бесконечными, можно задать широкий класс этих функций.

Линейные дискриминантные и разделяющие функции

Соотношения (Л15.5) и (Л15.6) приобретают наиболее простой вид, если в качестве функций $\left\{\varphi_j(\mathbf{x})\right\}$ выбрать линейные функции координат $\left(x_1, x_2, ..., x_n\right)$. Дискриминантные функции состояний в этом случае имеют вид

$$f^{(k)}(\mathbf{x})=\lambda_1^{(k)}x_1+\lambda_2^{(k)}x_2+...+\lambda_n^{(k)}x_n+\lambda_{n+1}^{(k)},\quad k=0,\ 1,\ ...,\ N \tag{Л15.7}$$

и называются *линейными дискриминантными функциями*, а коэффициенты $\lambda_j^{(k)}$ $\left(j=1,...,n+1\right)$ — *весовыми коэффициентами*, образующими *весовой вектор* $\boldsymbol{\lambda}^{(k)}=\left(\lambda_1^{(k)},\ ...,\ \lambda_{n+1}^{(k)}\right)$. Для удобства записи дискриминантных функций *n*-мерное пространство параметров дополняют компонентой x_{n+1}, причем соответствующую координату для векторов диагностируемых объектов полагают равной единице — $x_{n+1}\equiv 1$. Такое $\left(n+1\right)$-мерное пространство называется *дополненным пространством*. В этом пространстве с координатами $\left(x_1, x_2, ..., x_n, x_{n+1}\right)$ функция (Л15.7) записывается в виде скалярного произведения векторов

$$f^{(k)}\left(\mathbf{x}\right)=\boldsymbol{\lambda}^{(k)}\mathbf{x}=\sum_{j=1}^{n+1}\lambda_j^{(k)}x_j\,. \tag{Л15.8}$$

Если области диагнозов D_i и D_k соприкасаются, то разделяющая функция этих состояний определяется соотношением

$$f_{ik}(\mathbf{x}) = \boldsymbol{\lambda}^{(ik)}\mathbf{x}; \quad \boldsymbol{\lambda}^{(ik)} = \boldsymbol{\lambda}^{(i)} - \boldsymbol{\lambda}^{(k)} \tag{Л15.9}$$

и называется *линейной разделяющей функцией*. Разделяющей поверхностью двух состояний в дополненном пространстве является гиперплоскость, определяемая уравнением

$$f_{ik}(\mathbf{x}) = \boldsymbol{\lambda}^{(ik)}\mathbf{x} = \sum_{j=1}^{n+1} \lambda_j^{(ik)} x_j = 0. \tag{Л15.10}$$

Гиперплоскость проходит через начало координат дополненного пространства — точка $\mathbf{x} = 0$ удовлетворяет уравнению (Л15.10). Кроме того, так как для всех точек этой плоскости — векторов $\mathbf{x} = (x_1, x_2, ..., x_{n+1})$, скалярное произведение $\left(\boldsymbol{\lambda}^{(ik)}\mathbf{x}\right) = 0$, весовой вектор $\boldsymbol{\lambda}^{(ik)}$ нормален к гиперплоскости и определяет ее ориентацию в дополненном пространстве.

Построение линейных дискриминантных и разделяющих функций

Рассмотрим случай дихотомии, когда имеются два возможных состояния D_0 и D_1 — исправное и неисправное. Предположим области диагнозов D_0 и D_1 являются *линейно разделяемыми*, то есть существует весовой вектор $\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(01)}$, с помощью которого можно построить линейную разделяющую функцию (Л15.10) $f(\mathbf{x}) = f_{01}(\mathbf{x})$. Правило принятия решения о состоянии объекта в этом случае согласно (Л15.1) и (Л15.4) имеет вид

$$\left.\begin{aligned} &\mathbf{x} \in D_0, \text{если } f(\mathbf{x}) = \boldsymbol{\lambda}\mathbf{x} > 0; \\ &\mathbf{x} \in D_1, \text{если } f(\mathbf{x}) = \boldsymbol{\lambda}\mathbf{x} < 0. \end{aligned}\right\} \tag{Л15.11}$$

Так как разделяющая гиперплоскость в дополненном пространстве проходит через начало координат и однозначно определяется весовым вектором $\boldsymbol{\lambda}$, нормальным к этой плоскости, задача сводится к определению этого вектора.

Последний может быть определен с помощью обучающей выборки. Разделяющая функция для векторов обучающей выборки удовлетворяет условиям (Л15.11), второе из которых можно представить в виде

$$\mathbf{x} \in D_1, \text{ если } f(-\mathbf{x}) = \boldsymbol{\lambda}(-\mathbf{x}) = -\boldsymbol{\lambda}\mathbf{x} > 0.$$

Построим область $\overline{D}_1$, включающую векторы обучающей выборки из области D_1, у которых изменен знак. Таким образом, область $\overline{D}_1$ содержит векторы $-\mathbf{x}$ при условии, что $\mathbf{x} \in D_1$, и является симметричным отражением D_1 относительно начала координат. Объединим D_0 и

$\overline{D}_1$ в область $D = D_0 \bigcup \overline{D}_1$, для векторов которой разделяющая функция всегда положительна

$$f(\mathbf{x}) = \boldsymbol{\lambda}\mathbf{x} > 0\text{, если } \mathbf{x} \in D. \tag{Л15.12}$$

Последнее означает, что область D должна располагаться по одну сторону от разделяющей гиперплоскости.

Весовой вектор строится по следующему алгоритму. Из области D выбирают любой объект с известным диагнозом и вектором $\mathbf{x}_1$. В качестве первого приближения принимают $\boldsymbol{\lambda}_1 = \mathbf{x}_1$. Затем предъявляют следующий объект с вектором $\mathbf{x}_2$. Если $\boldsymbol{\lambda}_1\mathbf{x}_2 > 0$, то весовой вектор оставляют неизменным $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1$. Если же $\boldsymbol{\lambda}_1\mathbf{x}_2 < 0$, весовой вектор корректируется:

$$\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1 + \mathbf{x}_2.$$

Аналогично, если для любого последующего объекта с вектором $\mathbf{x}_{k+1}$ величина $\boldsymbol{\lambda}_k\mathbf{x}_{k+1} > 0$, то весовой вектор остается без изменения, если же $\boldsymbol{\lambda}_k\mathbf{x}_{k+1} < 0$, то принимают

$$\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_k + \mathbf{x}_{k+1}. \tag{Л15.13}$$

После того, как объекты обучающей выборки из области D исчерпаны, возвращаются к первому объекту, и процесс повторяется. Такая циклическая корректировка весового вектора продолжается до тех пор, пока все векторы области D не станут удовлетворять условию (Л15.11). Если области D_0 и D_1 линейно разделимы, рассмотренный алгоритм, называемый *правилом постоянного приращения*, приводит к определению необходимого весового вектора за конечное число шагов.

Пример Л15.1. Проиллюстрируем алгоритм построения весового вектора на простом и довольно искусственном примере построения линейной разделяющей функции двух областей, содержащих обучающие выборки из двух пар векторов $\mathbf{x}_1$, $\mathbf{x}_2$ — область D_0, и $\mathbf{x}_3$, $\mathbf{x}_4$ — область D_1, рис. Л15.2.

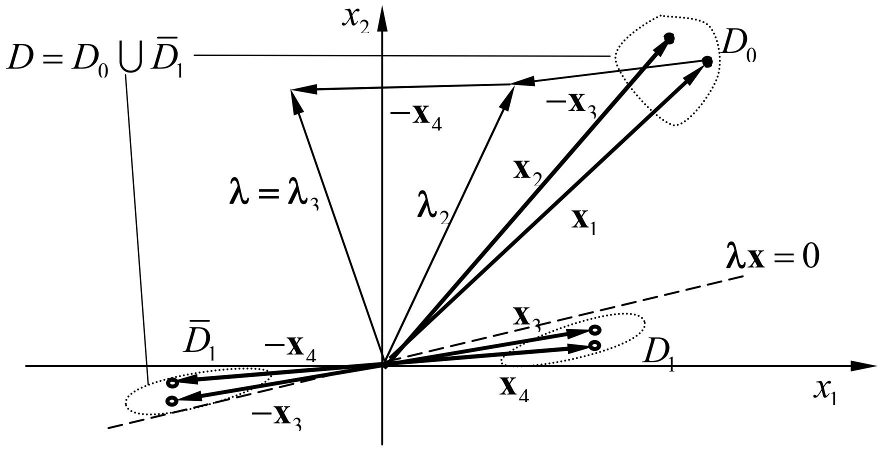

Рис. Л15.2. Иллюстрация алгоритма определения весового вектора и линейной разделяющей функции двух областей D_0 и D_1

Дополненное пространство с координатами (x_1, x_2, x_3) является трехмерным, и у всех векторов обучающих выборок третья координата $x_3 \equiv 1$. Рис. Л15.2 можно рассматривать как проекцию векторов на плоскость $x_3 = 1$ трехмерного дополненного пространства.

Построим область $\bar{D}_1$, включающую векторы области D_1, у которых изменен знак. В нашем случае область $\bar{D}_1$ содержит векторы $-\mathbf{x}_3$, $-\mathbf{x}_4$ и является симметричной области D_1 относительно начала координат. Далее объединим области D_0 и $\bar{D}_1$ в одну D, в которой содержится четыре вектора: $\mathbf{x}_1$, $\mathbf{x}_2$, $-\mathbf{x}_3$ и $-\mathbf{x}_4$.

В первом приближении в качестве весового вектора выберем $\boldsymbol{\lambda}_1 = \mathbf{x}_1$. Для второго вектора $\mathbf{x}_2$ проверяем условие $\boldsymbol{\lambda}_1\mathbf{x}_2 > 0$. Знак скалярного произведения векторов определяется знаком косинуса угла между этими векторами. Так как угол между векторами $\boldsymbol{\lambda}_1$ и $\mathbf{x}_2$ острый (рис. Л15.2), то скалярное произведение этих векторов положительно. Поэтому весовой вектор оставляем неизменным.

Далее из области D выбираем следующий вектор $-\mathbf{x}_3$ и проверяем условие $\boldsymbol{\lambda}_1 \cdot (-\mathbf{x}_3) > 0$. Угол между векторами $\boldsymbol{\lambda}_1$ и $-\mathbf{x}_3$ тупой, (рис. Л15.2), поэтому скалярное произведение $\boldsymbol{\lambda}_1 \cdot (-\mathbf{x}_3) < 0$. Проводим корректировку весового вектора. Во втором приближении он будет равен $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1 + (-\mathbf{x}_3)$, (рис. Л15.2).

Выбираем последний вектор $-\mathbf{x}_4$ из области D и проверяем условие $\boldsymbol{\lambda}_2 \cdot \left(-\mathbf{x}_4\right) > 0$. Поскольку угол между векторами $\boldsymbol{\lambda}_2$ и $-\mathbf{x}_4$ тупой (рис. Л15.2), скалярное произведение этих векторов отрицательно. Проводим очередную коррекцию весового вектора $\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_2 + \left(-\mathbf{x}_4\right)$ и проверяем знак скалярного произведения вектора $\boldsymbol{\lambda}_3$ с векторами области D: $\mathbf{x}_1$, $\mathbf{x}_2$, $-\mathbf{x}_3$ и $-\mathbf{x}_4$. Нетрудно видеть (рис. Л15.2), что все скалярные произведения положительные, так как углы между $\boldsymbol{\lambda}_3$ и всеми указанными векторами меньше 90°. Таким образом, весовой вектор равен $\boldsymbol{\lambda} = \boldsymbol{\lambda}_3$. Уравнение разделяющей плоскости, к которой нормален весовой вектор, имеет вид $\boldsymbol{\lambda}\mathbf{x} = 0$. Ее сечение с плоскостью $x_3 = 1$ дополненного пространства показано на рис. Л15.2 штриховой линией.

Алгоритм построения линейных дискриминантных функций при диагностике многих состояний во многом схож с рассмотренным выше и заключается в следующем. Предположим, что в i–ом приближении известен набор весовых векторов $\left\{\boldsymbol{\lambda}_i^{(k)}\right\}$ $\left(k = 0, 1, ..., N\right)$, и для распознавания предъявляется объект из обучающей выборки, характеризуемый вектором $\mathbf{x}_{i+1}$, с диагнозом D_j $\left(\mathbf{x}_{i+1} \in D_j\right)$. Корректировка весовых векторов проводится по следующему правилу:

весовые векторы не меняются, если справедливы неравенства

$$\boldsymbol{\lambda}_i^{(k)}\mathbf{x}_{i+1} > \boldsymbol{\lambda}_i^{(j)}\mathbf{x}_{i+1}, \;\; j = 0, ..., N; \;\; k \neq j, \qquad \text{(Л15.14)}$$

если для некоторых $j = n_1, n_2, ..., n_m$ $\boldsymbol{\lambda}_i^{(k)}\mathbf{x}_{i+1} < \boldsymbol{\lambda}_i^{(j)}\mathbf{x}_{i+1}$, то принимается

$$\boldsymbol{\lambda}_{i+1}^{(k)} = \boldsymbol{\lambda}_i^{(k)} + \mathbf{x}_{i+1}; \;\; \boldsymbol{\lambda}_{i+1}^{(j)} = \boldsymbol{\lambda}_i^{(j)} - \mathbf{x}_{i+1}; \;\; j = n_1, \, n_2, \, ..., \, n_m; \;\; k \neq j\,. \qquad \text{(Л15.15)}$$

Для корректировки используется бесконечная последовательность объектов, образуемая циклическим повторением объектов обучающей выборки. Корректировка прекращается, когда для всех векторов обучающей выборки будут выполнены неравенства (Л15.14).

Изложенные алгоритмы в зависимости от начального значения весового вектора и порядка предъявления объектов из обучающей выборки приводят к различным весовым векторам. Таким образом, в общем случае существует множество весовых векторов, решающих задачу постановки диагноза.