Лекция 6

Класификация с линейна обучаваща машина

Обучаващата машина може да бъде дефинирана като "устройство, чиито действия са повлияни от миналия опит" [1]. ***Линейната обучаваща машина*** (ЛОМ) е сравнително прост алгоритъм, който се побира в тази дефиниция, и чийто алгоритъм няма нищо общо с обучението, което претърпяват живите интелигентни организми. Той е подобен на изкуствените невронни мрежи, които са нелинейни по характера на своя алгоритъм – вижте лекция 11.

В най-разпространените химически приложения ЛОМ представлява съвкупността от един N-мерен вектор **W** и правилата за промяната на неговите координати w_n, наречени тегла [1,2]. Тази промяна се нарича ***обучение*** и се осъществява с помощта на обучаваща извадка, в която образите (обектите) са предварително разпределени в два класа. Всеки образ от обучаващата извадка се пропуска като вход в ЛОМ и се определя неговото класово разпределение: при грешно класифициране се извършва корекция на коефициентите на обучаващия вектор и обучението продължава със следващите образи. То спира при достигане на 100% ***разпознавателна способност*** или някакъв брой предварително определени преминавания през цялата обучителна извадка: едно преминаване се нарича ***цикъл***, ***сесия*** или ***епоха*** (session, epoch). Резултатите от това обучение могат да се проверят с помощта на тестваща извадка, с която може да се изчисли така наречената ***предсказваща способност*** – това е процентът на правилно класифицираните образи от обучаващата извадка: тук задължително се предполага, че това са различни образи от образите в обучаващата извадка.

Чрез поредица от двоични класификации ЛОМ може да се използва при мултикласификационни проблеми, но реално алгоритъмът за обучение е за разделяне на образите в два класа. За извършване на това обучение всички налични N-мерни образи с известни класове се превръщат в (N+1)-мерни като се прибавя (N+1) координата към тях, чиято стойност е еднаква за всички образи. След това те се разпределят случайно между двете извадки – обучаващата и тестващата. Обучаващата извадка се “разбърква”, т.е. обучението на ЛОМ се осъществява в случаен ред на срещане на различните образи, а не, например, първо с образите от единия клас, а после с тези от другия.

6.1. Алгоритъм. Нека за двата класа, 0 и 1, поставим следните изисквания за големината на скаларното произведение между тегловния вектор **W** и съответния образ $\mathbf{X_k}$.

$$\vec{X}_K \cdot \overrightarrow{W} > 0; \quad \vec{X}_K \in 1 \text{ и } \vec{X}_K \cdot \vec{W} < 0; \quad \vec{X}_K \in 0 \tag{6.1}$$

Нека в процеса на обучение за даден образ се получава грешно скаларно произведение

$$\vec{X}_K \cdot \vec{W} = S \tag{6.2}$$

т.е. ако обектът е от клас 1, S < 0, или ако обектът е от клас 0, то S > 0. Тогава корекцията на координатите на тегловния вектор **W** са следните (с прим е означен новият тегловен вектор):

$$\vec{W}' = \vec{W} + a\vec{X}_K \tag{6.3}$$

Като искаме така да променим тегловния вектор, че скаралното произведение да промени своята стойност в точно противоположната:

$$\vec{X}_K \cdot \vec{W}' = -S \tag{6.4}$$

Замествайки уравнения (2) и (3) в уравнение (4) получаваме

$$\vec{X}_K \cdot \vec{W}' = -S = \vec{W} \cdot \vec{X}_K + a\vec{X}_K \cdot \vec{X}_K = S + a\vec{X}_K \cdot \vec{X}_K$$

или

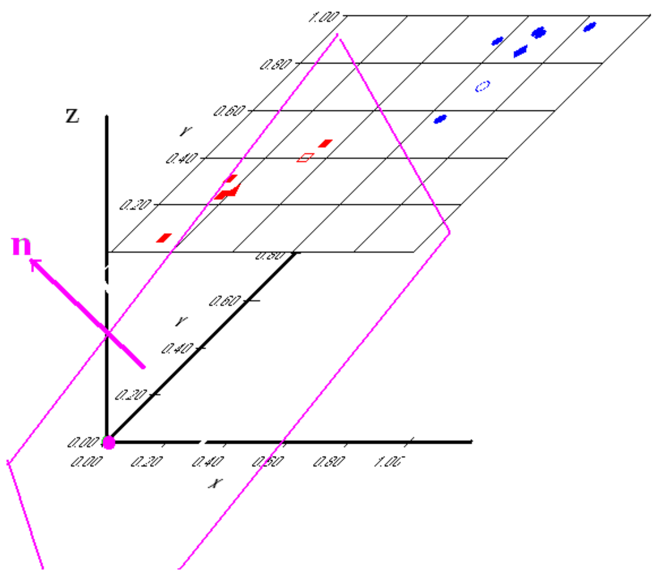
$$a = \frac{-2S}{\vec{X}_K \cdot \vec{X}_K} \qquad (6.5)$$

Обучението започва със случайни стойности на тегловния вектор **W**. В литературата ([2] и цитатите там) са описани изследвания с начални стойности w_n между -1 и +1, както и обучения, стартирали при всички координати на **W**, равни на единица или всички, равни на минус единица.

6.2. Разделяща хипер-равнина. Векторът **W** определя една хипер-равнина в (N+1)-мерното пространство, която има уравнение

$$\vec{X} \cdot \vec{W} = 0 \qquad (6.6)$$

което показва, че тази хиперравнина минава през началото на координатната система. Векторът **W** е перпендикулярен на тази равнина. На фигура 6.1 е показана тази равнина в тримерното пространство – образите са двумерни, но както бе споменато по-горе, към тях е прибавена трета координата с една и съща стойност, която ги отмества в равнина, успоредна на равнината xOy. На фигурата е показан нормалният вектор **n** към равнината – при нашите изисквания от уравнение (6.1), векторът **W** ще сочи в обратна посока на **n**. Освен това не е необходимо дължината на **W** да е единица, както е дължината на **n**. От фигурата се вижда и необходимостта да се добави допълнителна (в случая трета) координата – тя довежда до това, че разделящата равнина минава през началото на координатната система и затова могат да се формулират условията в уравнения (6.1) по този начин; вижда се, че правата която е сечение между двете равнини и която на практика разделя образите от двата класа не минава в общия случай през началото на координатната система. Илюстрацията е за двумерни образи, които с допълнителната трета координата се изобразяват като точки в тримерното пространство.



Фигура 6.1. Разделящата равнина, която се получава при пълно обучение на девет образа от два класа. Тези от клас 0 са означени с квадрати, тези от клас 1 – с кръгове, техните центроиди, съответно с триъгълник и ромб, а двата непознати образа с празен квадрат и празен кръг.

6.3. Обучение. Както споменахме обучението, т.е. корекцията по уравнение (6.3) се извършва само за образите, които дават "неправилно" скаларно произведение. За образите, които се класифицират правилно съобразно изискванията в уравнения (6.1) не се извършват корекции. Ако при едно пълно преминаване през обучаващата извадка няма нужда от нито една корекция, то обучението е завършено. На практика това става, т.е. алгоритъмът е сходим за линейно разделени образи (такива образи, между които може да се прекара разделяща хиперравнина, при което образите от единия клас са от едната и страна, а тези от другия от друга и страна). Ако образите са линейно неразделими обучението се спира след предварително зададен брой сесии (епохи, цикли). В този случай е очевидно, че разпознаващата способност ще е различна от 100%, и тъй като предсказващата способност е винаги по-малка от нея, то и тя ще е по-малка от 100%.

Според класификациите, дадени в предишната лекция, методът е непараметричен, изисква предварително знание за класовете, в които са съответните образи и без съмнение резултатът от прилагането му е класифициране на образите по класове.

Литература

1. K. Varmuza; *Chemometrics*. Springer Verlag, Berlin, 1980.
2. П. Джурс, Т. Айзенауэр; *Распознавание образов в химии*. Мир, Москва, 1977.

Задачи

Задача 6.1. Припомнете си уравнението на права в равнината. Колко коефициента има в него?

Задача 6.2. Припомнете си уравнението на равнина в тримерното пространство. Колко коефициента има в него?

Задача 6.3. Нека образите да са двумерни, а след прибавяне на трета координата към тях с постоянна стойност, $x_3 = 1.0$, те стават тримерни. Използвайте уравнение (6.6) и запишете уравнението на права в правоъгълната координатна система x_1Ox_2.

Задача 6.4. Изчислете скаларното произведение между векторите (образите):

$$X_1 = (1.1, -1.0, 2.2) \text{ и } X_2 = (-2.0, -1.2, 2.0).$$

Задача 6.5. Изчислете големината на първия вектор от задача 6.4.

Задача 6.6. Умножете четиримерния образ X = (1.2, -1.1, 0.1, 0.0) по -2.

Задача 6.7. Уравнението на равнина се извежда със скаларно произведение. Нека с $\mathbf{D}$ означим разстоянието[1] на равнината от началната точка на координатната система (0, 0, 0), с $\mathbf{n}$ - единичен нормален вектор[2] към равнината, а с $\mathbf{r}$ - вектор от началото към произволна точка на равнината. И

[1] Разстояние от точка до равнина е дължината на отсечката, която лежи на права, перпендикулярна към равнината и е между точката и точката на пресичане на правата с равнината.

[2] Разбира се свободен вектор – т.е. който има определени посока и големина, но не и точка на приложение.

точката и самият вектор **r** се дават с координати (x, y, z). Тогава е изпълнено:

nr = **D**.

a) Напишете скаларното произведение **nr**.

b) Преобразувайте го в уравнение на равнина – това, което сте записали в задача 6.2.

c) Уравнението на равнина също се дава с **nr** = **D**, но **r** има само две координати – **r = (x, y)**. **Като** напишете скаларното произведение **nr** преобразувайте **nr** = **D** в уравнение на права.