ОБУЧАЕМЫЕ КЛАССИФИКАТОРЫ. ДЕТЕРМИНИСТСКИЙ ПОДХОД

Общие свойства линейных дискриминантных функций в детерминистской постановке

Постникова Ольга гр.3341

Общие свойства линейных дискриминантных функций в детерминистской постановке

- **Здесь рассматривается задача классификации данных, заданных в виде конечных наборов многомерных векторов.**
- **Данный подход основан на нахождении линейных дискриминантных функций:**
- **$d(\overline{X}) = \overline{W}^T \overline{X} + W_{N+1} > 0$ (или $\leq$)**

- **Мы имеем следующее:**
- **$\overline{X} \in X1 \quad X1 = \{X_i\}_{i=1 \dots N_1}$**
- **$\overline{X} \in X2 \quad X2 = \{X_j\}_{j=1 \dots N_2}$**
- Общий объем выборки: $N = N_1 + N_2$
- **Задача заключается в нахождении решающей функции, которая удовлетворяет N линейным неравенствам, при условии N > n:**

$$(*)\begin{cases} \overline{W}^T \overline{X} + W_{n+1} > 0, \overline{X} \in X_1 \\ \overline{W}^T \overline{X} + W_{n+1} < 0, \overline{X} \in X_2 \end{cases} N > n$$

- Находим некоторую решающую функцию d(X), которая удовлетворяет неравенствам (*) и задает некоторую дихотомию, то есть разделение исходного пространства на два полупространства. Возникает вопрос можно ли решить данную систему неравенств. Возникает понятие разделяющей мощности решающего правила – это число возможных способов классификации данного объекта, которые допускаются с данной функцией.
- Можно рассмотреть количество линейных возможных дихотомий для N точек в линейном пространстве n. При этом каждая линейная решающая функция задает две дихотомии (так как нумерация классов может быть 1-2 или наоборот 2-1)
- Стоит задача разбиения точек в n-мерном пространстве с помощью (n-1) - мерной гиперплоскости.
- Общее возможных дихотомий для N точек равно 2N – это все
- возможные классификации: 2N = $\sum_{i-0}^{N} C_N^i$

- Оказывается, что не все возможные классификации могут быть заданы линейно. На рисунке представлены 4 точки , которые могут быть разделены с помощью 7 гиперплоскостей (в двумерном пространстве – просто линиями)
- Однако существуют дихотомии, которые не могут быть реализованы линейно

· x2 Линейно не могут быть заданы:

- ________________ I класс (x2, x4)
- x1 · · x3 II класс (x1, x3)
- · x4
- N = 4 Q = 24 = 16 QP = 16 – 2 = 14
- Формула, которая задает возможное количество классификаций (дихотомий), реализуемых линейно для N объектов, размерность пространства n:

$$D(N,n)=\begin{cases} 2\sum_{k=0}^{n} C_{N-1}^{k} & N>n \\ 2^{N} & N\le n \end{cases}$$

- Эта формула имеет место только тогда, когда точки объекта расположено “хорошо”. Это означает, что ни одна из точек группы, состоящей из (n+1) точки, не лежит в подпространстве размерности (n-1).

Пример

- Расчет количества возможных линейных дихотомий для N точек в n-мерном пространстве:

N \ n	1	2	3	4	5	6
1	2	2	2	2	2	2
10	20	92	260	512	764	932
200	400	39100	2627200	129109702		

- С ростом размерности число возможных дихотомий резко возрастает.
- Рассмотрим использование обобщенных линейных дискриминантных функций, полученных с помощью нелинейного преобразования исходного n-мерного пространства в пространство размерности $k>n$
- $d(X) = f1(x)W1 + f2(x)W2 + \ldots + fk(x)Wk + Wk+1$, где $k > n$

- Мы можем построить некие функции от х, путем некоего нелинейного преобразования и соответственно мы можем повысить размерность пространства и искать решение уже там.

$$\overline{\mathrm{Y}} = \begin{Bmatrix} f_1(\overline{X}) \\ \dots \\ f_k(\overline{X}) \end{Bmatrix} \leftarrow X = \begin{Bmatrix} X_1 \\ \dots \\ X_{k1} \end{Bmatrix}$$

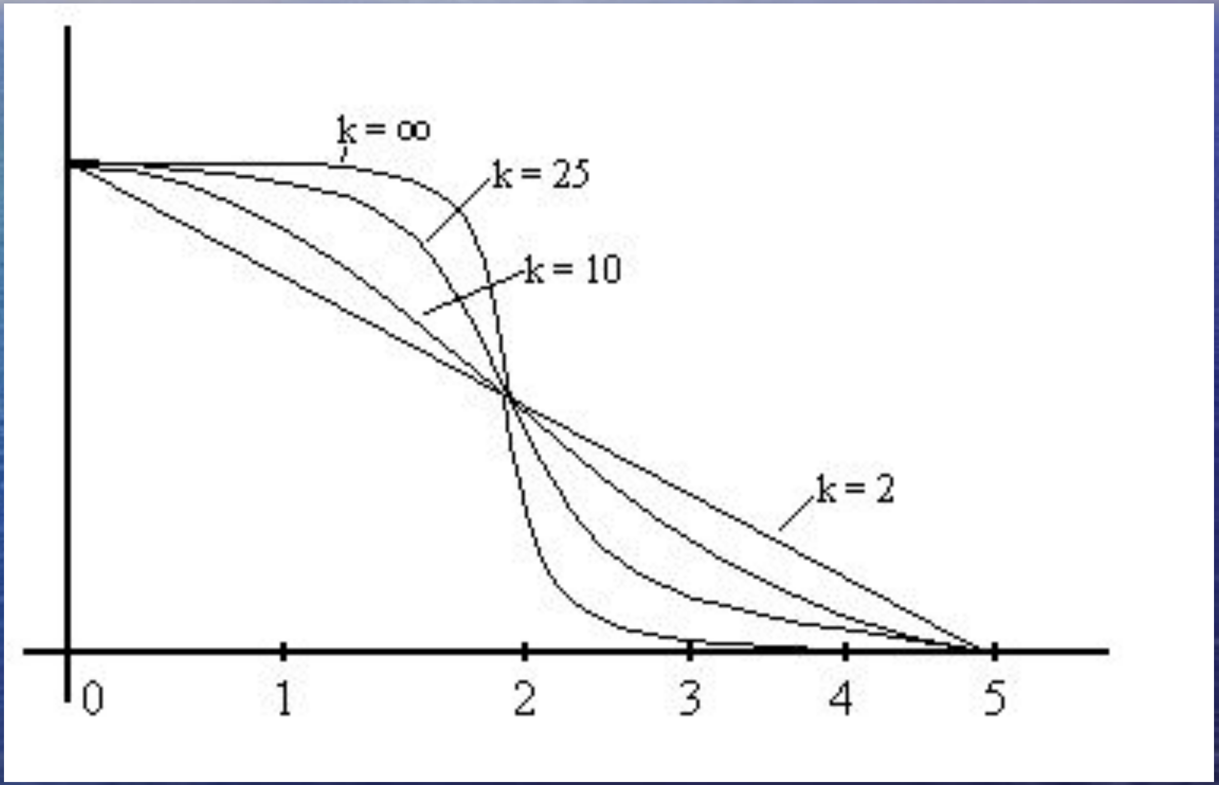
- Можно ввести понятие вероятность получения линейной дихотомии – это функция PN,K – вероятность того, что данная дихотомия будет реализована с помощью линейной функции.

$$\mathrm{N,K} = \frac{\mathrm{D(N,k)}}{2^{\mathrm{N}}} = \begin{Bmatrix} 2^{1-k} \sum_{\mathrm{j=0}}^{\mathrm{k}} \mathrm{C}_{\mathrm{N-1}}^{\mathrm{j}} & N > k \\ 1 & N \le k \end{Bmatrix}$$

• Как ведет себя данная функция?

- Определим параметр λ: $N = \lambda(k + 1)$
- Если ввести такой параметр, то получим, если k – обобщенная размерность, график зависимости

$P(N, \lambda)$



- Это зависимость вероятности получения линейной разделимости N точек при размерности пространства k

- При $\lambda < 2$ вероятность близка к единице.
- При $N < 2(k + 1)$ вероятность достаточно близка к единице.
- Величина $C_k = 2(k+1)$ называется мощностью соответствующая линейной решающей функции.
- Чем больше размерность, тем больше мощность решающей функции.
- Можно показать, что для исходного пространства с размерностью $\dim X = n$ мощность C_k для обобщенных линейных решающих функций определяется следующим образом:
- гиперплоскость – $C_k = 2(n+1)$;
- гиперсфера – $C_k = 2(n+2)$;
- поверхность второго порядка: $C_k = (n+1)(n+2)$
- полиномиальная поверхность порядка r: $C_k = 2C^r_{n+r}$

Персептронный алгоритм получения линейных решающих правил

Персептронный алгоритм получения линейных решающих правил

- Простейший методы получения линейных решающих функций на основе персептронных алгоритма обучения основывается на рекуррентном построении решающего правила путем коррекции ошибок.
- Требуется найти $\overline{W}^T, W_{n+1}$ для построения решающего правила

$$d(\overline{X}) = \overline{W}^T \overline{X} + W_{n+1}$$

на основе использования конечных обучающих выборок.

- Введем понятие расширенных векторов . Перейдем от размерности n к n+1 следующим образом:
- $$\widetilde{W} = \begin{pmatrix} W_1 \\ \dots \\ W_n \\ W_{n+1} \end{pmatrix} \qquad \widetilde{X} = \begin{pmatrix} X_1 \\ \dots \\ X_n \\ 1 \end{pmatrix}$$

- Тогда наша система неравенств сводится к более простой задаче:
- $(*)\ d(X) = \tilde{W}^T \tilde{X} > 0$ (или <) $x \in \mathbf{X}_1$ ($x \in \mathbf{X}_2$)
- Персептронный алгоритм основан на последовательном просмотре обучающей выборки:
- $X_1, \dots X_{N1} \dots X_N$
- $\mathbf{X}_1 \qquad \mathbf{X}_2$
- $N = N_1 + N_2$
- Процесс обучения заключается в том, что мы циклически просматриваем выборку и подставляем получаемое значение в W в (*), и на каждом шаге просмотра производим или не производим коррекцию весового вектора.
- 1. $\tilde{W}_{n+1} = \tilde{W}_n$, если $\begin{cases} x \in X_1, \tilde{W}_n^T \tilde{X} > 0 \\ x \in X_2, \tilde{W}_n^T \tilde{X} < 0 \end{cases}$
- В этом случае получен правильный ответ при классификации текущего вектора $\tilde{X}$

- 2. $\widetilde{W}_{n+1} = \widetilde{W}_n + C\overline{X}$, если $\widetilde{W}_n^T \widetilde{X} < 0,\ x \in X_1$
- $\widetilde{W}_{n+1} = \widetilde{W}_n - C\overline{X}$, если $\widetilde{W}_n^T \widetilde{X} > 0,\ x \in X_2$

Этот случай соответствует ошибочной классификации и соответственно производится коррекция весового вектора (должно быть C>0)

- Эта процедура и является процедурой обучения персептронного типа.
- Пусть мы имеем величину весового вектора после коррекции:

$$\widetilde{W}_{n+1} = \widetilde{W}_n + C\widetilde{X}$$

- Подставим новый весовой вектор в выражение для решающей функции:

$$\widetilde{W}_{n+1}^T \widetilde{X} = \left(\widetilde{W}_n + C\widetilde{X}\right)^T \widetilde{X} = \widetilde{W}_n^T \widetilde{X} + C\widetilde{X}^T \widetilde{X} = \widetilde{W}_n^T \widetilde{X} + C \| \widetilde{X} \|^2$$

- Видно, что значение весовой функции увеличилось на положительную
- величину $C \| \widetilde{X} \|^2$, то есть мы продвинулись к правильному решению.

- Показано, что если решение существует, то алгоритм сходится за конечное число шагов.
- Различные варианты выбора коэффициента C позволяют улучшить данный алгоритм:
- 1. C – константа . Скорость сходимости может быть мала.
- 2. $C = C_n = \mathrm{var}(n)$
- Попробуем менять C на каждом шагу так .чтобы сразу получить на текущем векторе правильное решение. Здесь можно использовать такой выбор
- $$C\widetilde{W}_n^T \widetilde{X} + C \|\widetilde{X}\|^2 > 0$$, отсюда следует $$C_n > \frac{|W_n^T X|}{\|\widetilde{X}\|^2}$$

- **Вывод:**

- Рассмотренный алгоритм появился на основе интуитивных соображений при разработке моделей работы головного мозга человека при решении задач обучения.