

10. Вероятностни и статистически методи за разпознаване. Подход на Бейс. Гаусово разпределение и разстояние по Мехаланобис. Методи за разпознаване по разстояние.

(Виж материалите страница ПР_Част1 41 – 44, <http://www2.it.lut.fi/kurssit/04-05/010586001/lectures.html> - лекции 1 и 2)

Вероятностна функция

$p : W \rightarrow [0,1]$, където W е n -мерно Евклидово пространство.

Дадени са вероятностни класове $C_1, C_2, \dots, C_n$ с вероятности:

$P(C_1) > P(C_2) > \dots > P(C_n)$

Условия дадена функция да е вероятностна:

1. $p(x) \in [0,1]$
2. $p(W) = 1$
3. Ако $A, B \in W$ и $A \cap B = \emptyset$, то $p(A+B) = p(A) + p(B)$
4. Съвместно събитие – $p(A, B) = p(A) + p(B) - p(A \cap B)$.

Събитие $A - p(x), x \in W$

Събитие $B -$ Избор на класа на $x - x \rightarrow \text{Recognizer} \rightarrow i \Rightarrow x \in C_i$

$p(x, C_i) = p(C_i|x) * p(x) = p(x|C_i) * p(C_i)$

$$p(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)}$$

където $p(C_i|x)$ – апостериорната вероятност, образът X да попадне в класа C_i ,

$p(x|C_i)$ е вероятността за появата на образ X , при условие че разпознаването е показало C_i

$p(C_i)$ е априорната вероятност за появата на образ от този клас

Подход на Бейс.

Класификация при максимална апостериорна вероятност

$y = \text{Argmax } p(C_i|x)$ (Избираме класа с максимална апостериорна вероятност)

$p(C_y|x) \geq p(C_i|x)$, за всяко $i=1, \dots, n$.

Нека $g_i(x) = p(C_i|x)$, тогава класификаторът има вида:

$$\begin{array}{l} g_1(x) \\ \rightarrow g_2(x) \rightarrow \max? \rightarrow y \\ \dots \\ g_n(x) \end{array}$$

Класификация при минимална грешка

Правило на Бейс:

$$p(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)}$$

$$p(x) = \sum_{i=1}^n p(x|C_i)p(C_i)$$

Грешката при класификация на два класа би изглеждала:

$$p(error|x) = \frac{P(C1|x) - \text{при класифициран } C2}{P(C2|x) - \text{при класифициран } C1}$$

Средната грешка би изглеждала като $\hat{p}(error) = \int p(error|x)p(x)dx$

В контекста на подхода на Бейс грешката е $p(error|x) = \min\{p(C1|x), p(C2, x)\}$.

Вероятността за грешна класификация p_{error} при повече класове е:

$$p(error|x) = 1 - \sum_{x \in W} P(Ci|x)$$

което реално е вероятността x да не принадлежи на Ci . Необходимо е да споменем, че при несъвместимост на входа x , то и класовете също са несъвместими.

Класификация при минимален риск

Разделяме пространството на класове.

Нека всяко едно класификационно действие има цена - λ_{ki} ($x \in Ri$ (Регион от класа Ci), а правилния клас е Ck).

$$r_k = \sum_{i=1}^n \lambda_{ki} \int_{R_i}^{\sim} p(x|Ck)dx,$$

където интеграла е вероятността вектор от Ck да бъде класифициран като вектор от Ci .
Така за средния риск получаваме:

$$\hat{r} = \sum_{k=1}^n r_k p(Ck) = \sum_{i=1}^n \int_{R_i}^{\sim} \left(\sum_{k=1}^n \lambda_{ki} p(x|Ck) p(Ck) \right) dx.$$

Сега нека дефинираме li – като функцията на загубената стойност за класа Ci така:

$$l_i = \sum_{k=1}^n \lambda_{ki} p(x|Ck) p(Ck).$$

Класификатора за минимален риск класифицира x в региона $Ri \leftrightarrow li < lj$ за всяко $i \neq j$.

Да разгледаме случая с два класа:

$$l_1 = \lambda_{11} p(x|C1) p(C1) + \lambda_{21} p(x|C2) p(C2)$$

$$l_2 = \lambda_{12} p(x|C1) p(C1) + \lambda_{22} p(x|C2) p(C2)$$

Правилото би било да се избере $C1$, ако $l1 < l2$.

$$(\lambda_{21} - \lambda_{22}) p(x|C2) p(C2) < (\lambda_{12} - \lambda_{11}) p(x|C1) p(C1)$$

$\lambda_{21} - \lambda_{22} > 0$ – обикновено.

$\Rightarrow$ Избери $C1$ ако $\frac{p(x|C1)}{p(x|C2)} > \frac{p(C2) (\lambda_{21} - \lambda_{22})}{p(C1) (\lambda_{12} - \lambda_{11})}$, където дясната страна се нарича коефициент на подобие.

Дискриминантни функции и разделящи повърхности

Дискриминантния анализ се занимава с границите между регионите Ri , а не с разпределенията вътре в тях. При дискриминантния анализ се знаят границите на класовете, но не се знае разпределението вътре в тях. Класовете са непресичащи се по между си.

Когато $p(Ci|x) - p(Cj|x) = 0$ имаме, че x е на граничната повърхност м/у Ci и Cj .

Дефинираме $gi(x) = f(p(Ci|x))$, където $f()$ е произволна монотонно растяща функция.

Тези функции $gi(x)$ се наричат дискриминантни функции.

От подхода на Бейс имаме – избери Ci , ако $gi(x) \geq gj(x)$, за всяко $i \neq j$.

Разделящата повърхност би се характеризирала с функцията:

$G_{ij}(x) = g_i(x) - g_j(x) = 0$ точн тогава, когато x е на границата между двата класа.

При дискриминантния анализ се правят много допускания това го прави по не точен от статистическите подходи.

Да допуснем, че класовете имат нормална плътност

Гаусово разпределение и разстояние на Мехаланобис.

Да предположим, че всеки един от класовете има нормално разпределение около някакъв център.

$N(\mu, \Sigma, x)$ – гаусиан.

$$p(x|C) = N(\mu, \Sigma, x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu-x)^T \Sigma^{-1}(\mu-x)} \text{ – за } n\text{-мерен гаусиан } x \in E^n.$$

μ – средна стойност на математическото очакване за $x(E[x])$.

Σ - ковариационна матрица, и $\Sigma = \|\sigma_{ij}\|_{n \times n}$ за нея имаме $\sigma_{ij}^2 = E[(\mu_i - x_i)(\mu_j - x_j)]$

Σ е само с положителни стойности и е симетрична.

σ_{kk} е дисперсията за x_k , σ_{km} са съответните ковариации, а ако x_k и x_m са независими, то $\sigma_{km} = 0$.

$$N(\mu, \delta, x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \text{ в двумерния случай.}$$

Да се опитаме да построим класификатор на Бейс.

За дискриминантна функция избираме натурален логаритъм.

$$g_i(x) = \ln(p(x|C_i)p(C_i)) = \ln p(x|C_i) + \ln p(C_i)$$

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln p(C_i)$$

Да построим функцията на граничната повърхност:

$G_{ij}(x) = g_i(x) - g_j(x)$, където $x \in E^2$ (1D е точка, 2D е крива).

$$G_{ij}(x) = 0 \iff g_i(x) = g_j(x)$$

$$(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + \ln |\Sigma_i| - \ln p(C_i) = (x - \mu_j)^T \Sigma_j^{-1}(x - \mu_j) + \ln |\Sigma_j| - \ln p(C_j)$$

1. Да приемем, че $\Sigma_i = \Sigma$, за всяко i .

Тогава, ако разпишем $(x - \mu_i)^T \Sigma^{-1}(x - \mu_i) = \Sigma^{-1} x^T x - 2 \Sigma^{-1} \mu_i^T x + \mu_i^T \Sigma^{-1} \mu_i$, така получаваме един и за 2-рата степен на x , че не зависи от C_i

$\Rightarrow$ Като махнем константите и 2-рата степен получаваме линейна функция:

$$g_i(x) = \Sigma^{-1} \mu_i^T x - \frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i + \ln p(C_i)$$

$$g_i(x) = w^T x + w_0$$

От геометрична гледна точка става въпрос за хипер равнина.

$$1.1 \Sigma = \sigma^2 I$$

$$g_i(x) = \frac{1}{\sigma^2} \mu_i^T x + w_0$$

Сега за $G_{ij}(x)$ имаме

$$G_{ij}(x) = w^T (x - x_0)$$

$$w^T = \mu_i - \mu_j$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \sigma^2 \ln\left(\frac{p(Ci)}{p(Cj)}\right) \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|^2}$$

⇒ граничната повърхност е равнина през x_0 и перпендикулярна на $\mu_i - \mu_j$, а при еднаква априорна вероятност за класовете имаме, че класификатора се свежда до намиране на минималното евклидово разстояние:

$$D_E = \|x - \mu_i\|$$

1.2 Σ - да не е диагонална

$$G_{ij}(x) = w^T(x - x_0)$$

$$w^T = \Sigma^{-1}(\mu_i - \mu_j)$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \sigma^2 \ln\left(\frac{p(Ci)}{p(Cj)}\right) \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|_{\Sigma^{-1}}^2}$$

Като $\|x\|_{\Sigma^{-1}} = \sqrt{x^T \Sigma^{-1} x}$ – което се нарича разстояние по Мехаланобис.

⇒ граничната повърхност е равнина през x_0 перпендикулярна на $\Sigma^{-1}(\mu_i - \mu_j)$, а при еднаква априорна вероятност за класовете имаме, че класификатора се свежда до намиране на минималното разстояние по Мехаланобис:

$$D_M = \sqrt{(x - \mu_i)^T \Sigma^{-1} (x - \mu_i)}$$

2. При различни Σ :

$$g_i(x) = x^T W_i x + w_i^T x + w_{i0}$$

$$W_i = -\frac{1}{2} \Sigma_i^{-1}$$

$$w_i = \Sigma_i^{-1} \mu_i$$

$$w_{i0} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln p(Ci)$$

Повърхностите са хипер квадрати и класовете на са просто свързани.

Приложението на разстоянието на Мехаланобис е за клъстаризация и класификация на обекти, използва се в дискриминантния анализ и в линейния дискриминантен анализ на Фишер.