

Задачи по байесовскому подходу к классификации

Денисов Д.М.

Исходные данные:

$$p(x|y = -1) = \frac{1}{\pi(1+x^2)} \sim \text{Cauchy}(0, 1),$$

$$p(x|y = +1) = \frac{1}{3}I_{[a,b]} \sim U(0, 3),$$

$$\lambda_- = 2, \lambda_+ = 1,$$

$$p(y = -1) = 0.4, p(y = +1) = 0.6.$$

1 Оптимальный байесовский классификатор

Имеем:

$$\begin{aligned} a^*(x) &= \operatorname{argmax}_y \lambda_y P(y) p(x|y) \\ &= \operatorname{argmax}_y \{ \lambda_- P(y = -1) p(x|y = -1), \lambda_+ P(y = +1) p(x|y = +1) \} \\ &= \operatorname{argmax}_y \left\{ \frac{0.8}{\pi(1+x^2)}, 0.2I_{[a,b]} \right\} \\ &= \operatorname{argmax}_y \{ f(x), g(x) \}. \end{aligned}$$

Графики распределений $f(x)$ и $g(x)$ представлены на рис. 1.
Из рис. 1 видно, что

$$g(x) \geq f(x) \iff x \in [x_0, 3].$$

Таким образом, оптимальный классификатор будет иметь вид

$$a^*(x) = \begin{cases} -1, & x \in (-\infty, x_0) \cup (3, +\infty) \\ +1, & x \in [x_0, 3] \end{cases}.$$

Координату x_0 найдём следующим образом:

$$\begin{aligned} f(x_0) = g(x_0) = 0.2 &\implies \frac{0.8}{\pi(1+x_0^2)} = 0.2 \implies \\ &\implies \pi(1+x_0^2) = 4 \implies x_0 = \sqrt{\frac{4}{\pi} - 1}. \end{aligned}$$

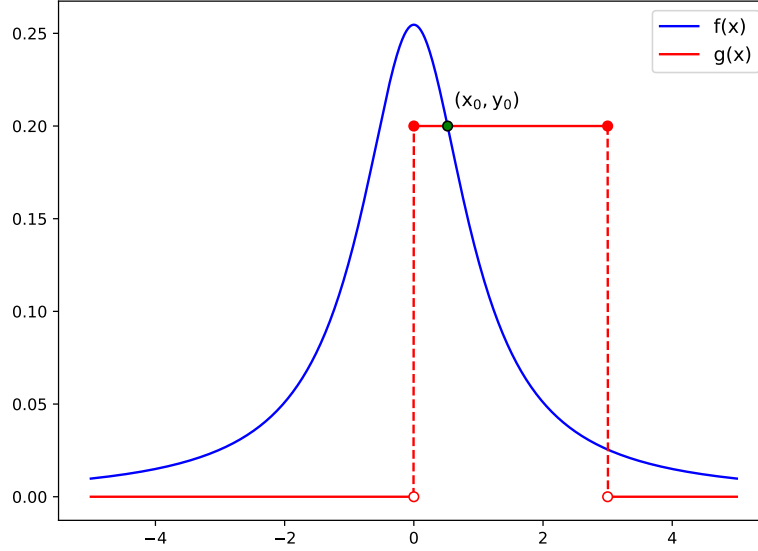


Рис. 1: Графики распределений

2 Оптимальный средний риск

Имеем:

$$\begin{aligned}
 R^* &= \iint L(a^*(x), y) p(x, y) dx dy \\
 &= \iint \lambda_y [a^*(x) \neq y] p(x|y) P(y) dx dy \\
 &= \iint \lambda_y (1 - [a^*(x) = y]) p(x|y) P(y) dx dy \\
 &= \int \min_y \lambda_y p(x|y) P(y) dx dy \\
 &= \int \min_y \{ \lambda_- p(x|y = -1) P(y = -1), \lambda_+ p(x|y = +1) P(y = +1) \} dx dy \\
 &= \int \min_y \{ f(x), g(x) \} dx dy.
 \end{aligned}$$

Аналогично пункту 1, получаем:

$$\begin{aligned}
 R^* &= \int_{-\infty}^{x_0} g(x) dx + \int_{x_0}^3 f(x) dx + \int_3^{+\infty} g(x) dx \\
 &= \int_0^{x_0} 0.2 dx + \int_{x_0}^3 \frac{0.8}{\pi(1+x^2)} dx \\
 &= 0.2x_0 + \frac{0.8}{\pi} (\arctan 3 - \arctan x_0) \approx 0.299958 \approx 0.3.
 \end{aligned}$$

3 Наивный байесовский классификатор

Найдём наивный байесовский классификатор так: