Задачи по байесовскому подходу к классификации

Денисов Д.М.

Исходные данные:

$$p(x|y=-1) = \frac{1}{\pi(1+x^2)} \sim Cauchy(0,1),$$
$$p(x|y=+1) = \frac{1}{3} I_{[a,b]} \sim U(0,3),$$
$$\lambda_- = 2,\ \lambda_+ = 1,$$
$$p(y=-1) = 0.4,\ p(y=+1) = 0.6.$$

1 Оптимальный байесовский классификатор

Имеем:

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

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:

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

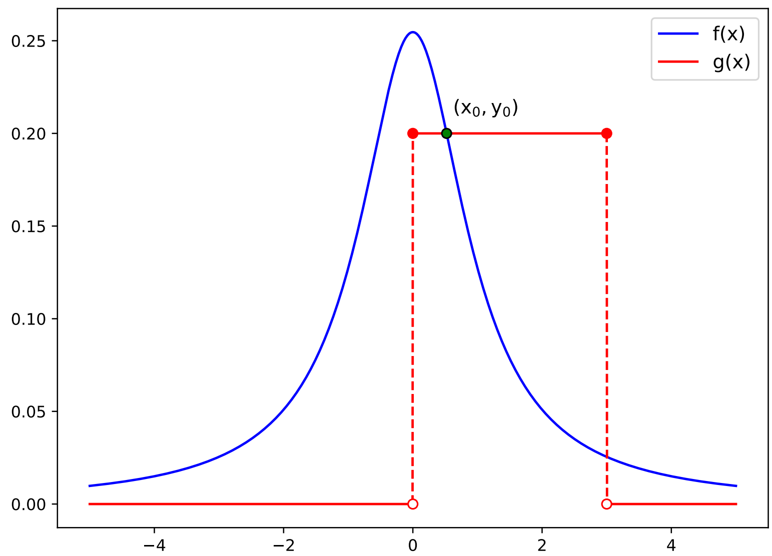

Рис. 1: Графики распределений

2 Оптимальный средний риск

Найдём оптимальный средний риск:

3 Наивный байесовский классификатор

Найдём наивный байесовский классификатор так: