Задача 1

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

$$p(x \mid y=1) = \text{Exp}(2) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Функция потерь: $L(a(x), y) = \lambda_y [a(x) \neq y]$,
где $\lambda_{y=1} = 2$, $\lambda_{y=-1} = 1$

$$P(y=-1) = 0{,}4 \Rightarrow P(y=1) = 0{,}6$$

Найдем оптимальный байесовский классификатор.

$$a(x) = \arg\max_y \lambda_y \, p(y \mid x) \text{ — ?}$$

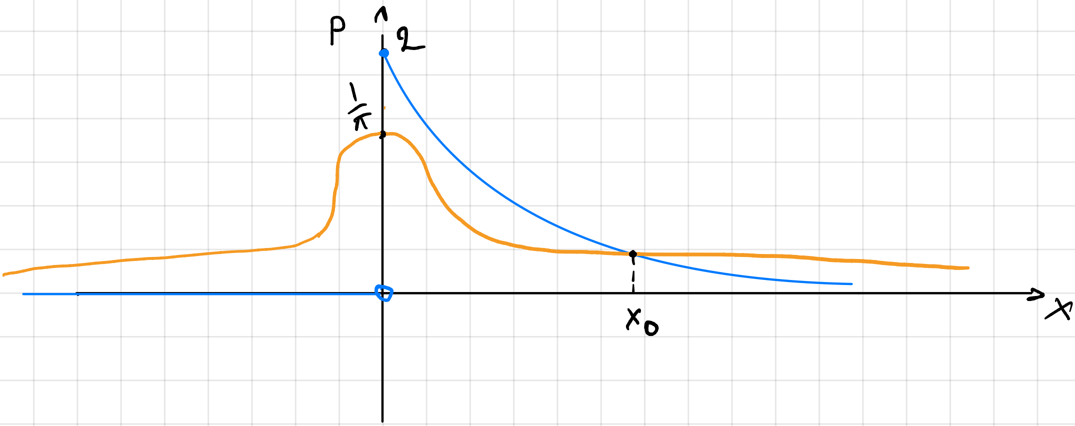
$$p(y \mid x) = \frac{P(y)\, p(x \mid y)}{P(x)} \text{ — по ф. Байеса.} \Rightarrow$$

$$\Rightarrow a(x) = \arg\max_y \lambda_y \frac{P(y)\, p(x \mid y)}{p(x)} =$$

$$= \arg\max_y \lambda_y P(y)\, p(x \mid y) =$$

$$= \arg\max_y \begin{cases} \lambda_{y=1} \cdot P(y=1)\, p(x \mid y=1) \\ \lambda_{y=-1} \cdot P(y=-1)\, p(x \mid y=-1) \end{cases} =$$

$$= \arg\max_y \begin{cases} 2 \cdot 0{,}6 \cdot \mathrm{Exp}(2) \\ 1 \cdot 0{,}4 \cdot \mathrm{Cauchy}(0,1) \end{cases}$$



По графику видно, что:

$p(x \mid y=1) > p(x \mid y=-1)$ при $x \in [0, x_0]$

$p(x \mid y=-1) > p(x \mid y=1)$ при $x \in (-\infty; 0) \cup (x_0; +\infty)$

Найдем x_0: $2 \cdot 0{,}6 \cdot 2 e^{-2x} = 0{,}4\, \frac{1}{\pi} \left(\frac{1}{1+x^2} \right)$

$x_0 \approx 2{,}436.$

Ответ:

$$a(x) = \begin{cases} 1, & x \in [0; 2{,}436] \\ -1, & x \in (-\infty; 0) \cup (2{,}436; +\infty) \end{cases}$$

Задача 2.

Найти R – средний риск.

Решение:

$$R(a) = \iint\limits_{X\,Y} \lambda_y L(a(x), y)\, p(x,y)\, dx dy =$$

$$= \int\limits_X \sum_y \lambda_y [a(x) \neq y]\, p(x|y)\, P(y)\, dx =$$

$$= \int\limits_X \sum_y \lambda_y \left(1 - [a(x) = y]\right) p(x|y)\, P(y)\, dx =$$

$$= \int\limits_X \sum_y \lambda_y\, p(x|y)\, P(y)\, dx - \int\limits_X \sum_y \lambda_y [a(x)=y]\, p(x|y) P(y) dx$$

$$= \int\limits_X \lambda_{y=1}\, p(x|y=1)\, P(y=1)\, dx + \int\limits_X \lambda_{y=-1}\, p(x|y=-1)\, P(y=-1)\, dx -$$

$$- \int\limits_X \lambda_{y=1} [a(x)=1]\, p(x|y=1)\, P(y=1) dx - \int\limits_X \lambda_{y=-1} [a(x)=-1]\, p(x|y=-1) P(y=-1)$$

$$= \int\limits_X 2 \cdot Exp(2) \cdot 0{,}6\, dx + \int\limits_X Cauch(0,1) \cdot 0{,}4\, dx -$$

$$- \int_0^{x_0} 2 \cdot Exp(2) \cdot 0{,}6\, dx - \int\limits_{(-\infty;0) \cup (x_0;+\infty)} 0{,}4 \cdot Cauch(0,1)\, dx =$$

$$= \int_0^{+\infty} 2\cdot 0{,}6\cdot 2e^{-2x}\,dx + \int_{-\infty}^{+\infty} \frac{0{,}4}{\pi}\left(\frac{1}{1+x^2}\right)dx -$$

$$- \int_0^{x_0} 2\cdot 0{,}6\cdot 2e^{-2x}\,dx - \frac{0{,}4}{\pi}\left[\int_{-\infty}^{0} \frac{1}{1+x^2}\,dx + \int_{x_0}^{+\infty} \frac{1}{1+x^2}\,dx\right] =$$

$$= 2{,}4\cdot\left(-\frac{1}{2}\left(e^{-\infty}-1\right)\right) + \frac{0{,}4}{\pi}\left(\operatorname{arctg}(\infty) - \operatorname{arctg}(-\infty)\right) +$$

$$+ 1{,}2\left(e^{-2x_0}-1\right) - \frac{0{,}4}{\pi}\left(\operatorname{arctg}(\infty) - \operatorname{arctg}(-\infty) - \operatorname{arctg}(x_0)\right) =$$

$$= 1{,}2 + 0{,}4 + 1{,}2\left(e^{-2x_0}-1\right) - \frac{0{,}4}{\pi}\left(\pi - \operatorname{arctg}(x_0)\right) =$$

$$= 1{,}6 + 1{,}2\left(e^{-2x_0}-1\right) - \frac{0{,}4}{\pi}\left(\pi - \operatorname{arctg}(x_0)\right)$$

Ответ:

$$R = 1{,}6 + 1{,}2\left(e^{-2x_0}-1\right) - \frac{0{,}4}{\pi}\left(\pi - \operatorname{arctg}(x_0)\right)$$

при $x_0 \approx 2{,}436$

$R \approx 0{,}160$

Задача 3.

Дано:

$L(a(x), y) = [a(x) \neq y]$,

данные из 'task2.csv'

task2.csv

	x_1	x_2	target
0	-0.027000	3.610953	-1
1	-0.556097	-0.015698	-1
2	-0.334942	2.280331	-1
3	0.430564	-3.189805	-1
4	-0.314852	-2.071801	-1

$x = (x_1, x_2)$

Найти наивный байесовский классификатор:

$a(x)$ — ?

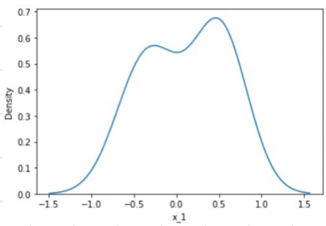
Решение: $a(x) = \arg\max\limits_{y} p(x|y)P(y)$

$p(y=1) = 0{,}6$

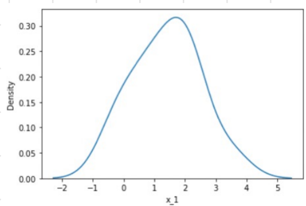
$p(y=-1) = 0{,}4$

$p(x|y) = p(x_1|y)\,p(x_2|y)$ (так как классиф-ор наивный)

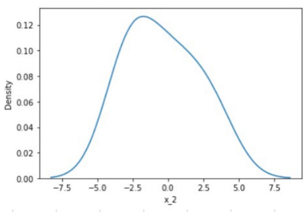
$p(x\_1 | y=-1)$



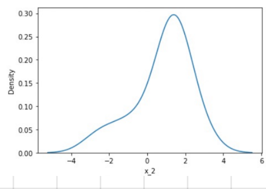
$p(x\_1 | y=1)$



$p(x\_2 | y=-1)$



$p(x\_2 | y=1)$



Будем считать, что x имеет нормальное распр.

$p(x\_1 \mid y=-1) \sim N(0.1,\ 0.2)$

$p(x\_2 \mid y=-1) \sim N(-0.4,\ 6.4)$

$p(x\_1 \mid y=1) \sim N(1.4,\ 1.2)$

$p(x\_2 \mid y=1) \sim N(0,8,\ 2.2)$

$$a(x) = \arg\max_y p(x \mid y)\, P(y) =$$

$$= \arg\max_y \begin{Bmatrix} p(x\_1 \mid y=1)\, p(x\_2 \mid y=1)\, P(y=1) \\ p(x\_1 \mid y=-1)\, p(x\_2 \mid y=-1)\, P(y=-1) \end{Bmatrix}$$

$$0{,}6 \cdot \frac{1}{\sqrt{2\pi\sigma_1^{+2}}} \exp\left(-\frac{(x_1-\mu_1^+)^2}{2\sigma_1^{+2}}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^{+2}}} \exp\left(-\frac{(x_2-\mu_2^+)}{2\sigma_2^{+2}}\right) \geqslant$$

$$\geqslant 0{,}4 \frac{1}{\sqrt{2\pi\sigma_1^{-2}}} \exp\left(-\frac{(x_1-\mu_1^-)^2}{2\sigma_1^{-2}}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^{-2}}} \exp\left(\frac{-(x_2-\mu_2^-)^2}{2\sigma_2^{-2}}\right)$$

$$0{,}72 \exp\left[-\frac{(x_1-1.4)^2}{2\cdot 1{,}2^2} + \frac{(x_1-0{,}1)^2}{2\cdot 0.2^2}\right] \geqslant \exp\left[-\frac{(x_2+0{,}4)^2}{2\cdot 6.4^2} + \frac{(x_2-0{,}8)^2}{2\cdot 2.2^2}\right]$$

$$0{,}72 \exp\left(\frac{(x_1-0{,}3)(x_1+0{,}2)}{0{,}115}\right) \geqslant \exp\left(\frac{(x_2-0{,}5)(x_2-1{,}4)}{396}\right)$$

Wolfram Alpha: где $y = x_2$, $x = x_1$

Real solutions

$x < -0.266574$,
$0.05\left(19 - 3\sqrt{17600\log(0.72 \times 2.71828^{8.69565(x-0.3)(x+0.2)}) + 9}\right) < y < 0.05\left(3\sqrt{17600\log(0.72 \times 2.71828^{8.69565(x-0.3)(x+0.2)}) + 9} + 19\right)$

$x > 0.366574$, $0.05\left(19 - 3\sqrt{17600\log(0.72 \times 2.71828^{8.69565(x-0.3)(x+0.2)}) + 9}\right) < y < 0.05\left(3\sqrt{17600\log(0.72 \times 2.71828^{8.69565(x-0.3)(x+0.2)}) + 9} + 19\right)$

Область B.

Ответ:

$$a(x) = \begin{cases} 1, & \text{при } (x_1, x_2) \in B \\ -1, & \text{при } (x_1, x_2) \notin B \end{cases}$$

$$= 2,4\left(-\frac{1}{2}e^{-2x}\Big|_0^{\infty}\right) + \frac{0,4}{\pi}\left(\operatorname{arctg}(\infty) - \operatorname{arctg}(\infty)\right) +$$

$$+ 1,2\; e^{-2x}\Big|_0^{x_0} - \frac{0,4}{\pi}\left(\operatorname{arctg}(0) - \operatorname{arctg}(-\infty)\right) -$$

$$- \frac{0,4}{\pi}\left(\operatorname{arctg}(\infty) - \operatorname{arctg}(x_0)\right) =$$

$$= 1,2 + 0,4 + 1,2\left(e^{-2x_0} - 1\right) - \frac{0,4}{\pi}\left(\pi - \operatorname{arctg}(x_0)\right)$$