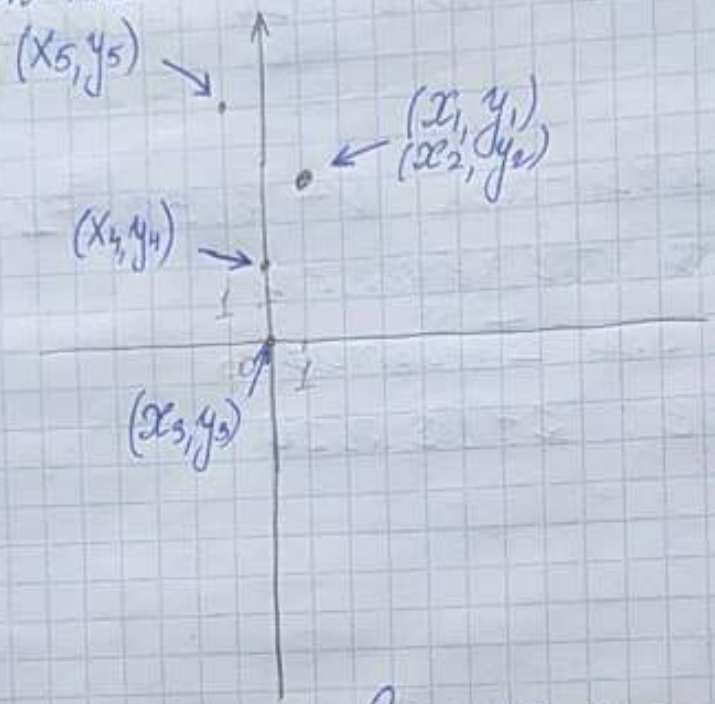
Задача 3.

x	1	1	0	0	-1
y	4	4	0	2	-6

– обучающая выборка

1) Изобразить точки



2) МНК для модели $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ – аппроксим. параб. функцией

$$X = \begin{pmatrix} 1 & 1 & x_1^2 \\ 1 & 1 & x_2^2 \\ 1 & 0 & x_3^2 \\ 1 & 0 & x_4^2 \\ 1 & -1 & x_5^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{5\times 3}$$

$$y = \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} \in \mathbb{R}^{5\times 1} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^{3\times 1}$$

Цель: методом наим. квадратов минимизировать $\|y - X\beta\|_2^2 \to \min$

$$X^T X \beta = X^T y \mid \cdot (X^T X)^{-1}$$

$$\beta = (X^T X)^{-1} \cdot X^T \cdot y$$

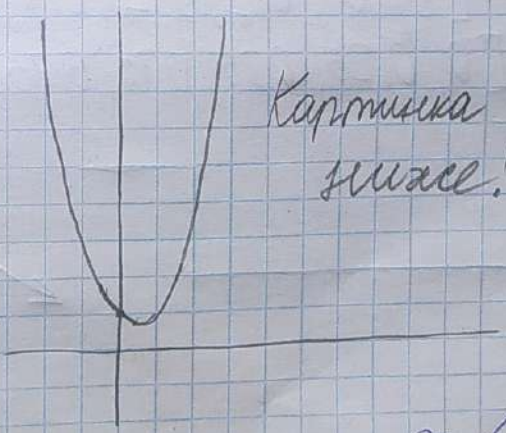
$$X^T \cdot X = \underset{5\times 3}{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix}} \cdot \underset{3\times 3}{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}} = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix}$$

$$X^T \cdot y = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}.$$ Для нахождения β необязательно считать $(X^T X)^{-1}$.

$$\begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix} \xrightarrow{\text{Метод Гаусса}} \beta = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$$

$$f(x) = 1 - x + 4x^2$$

*

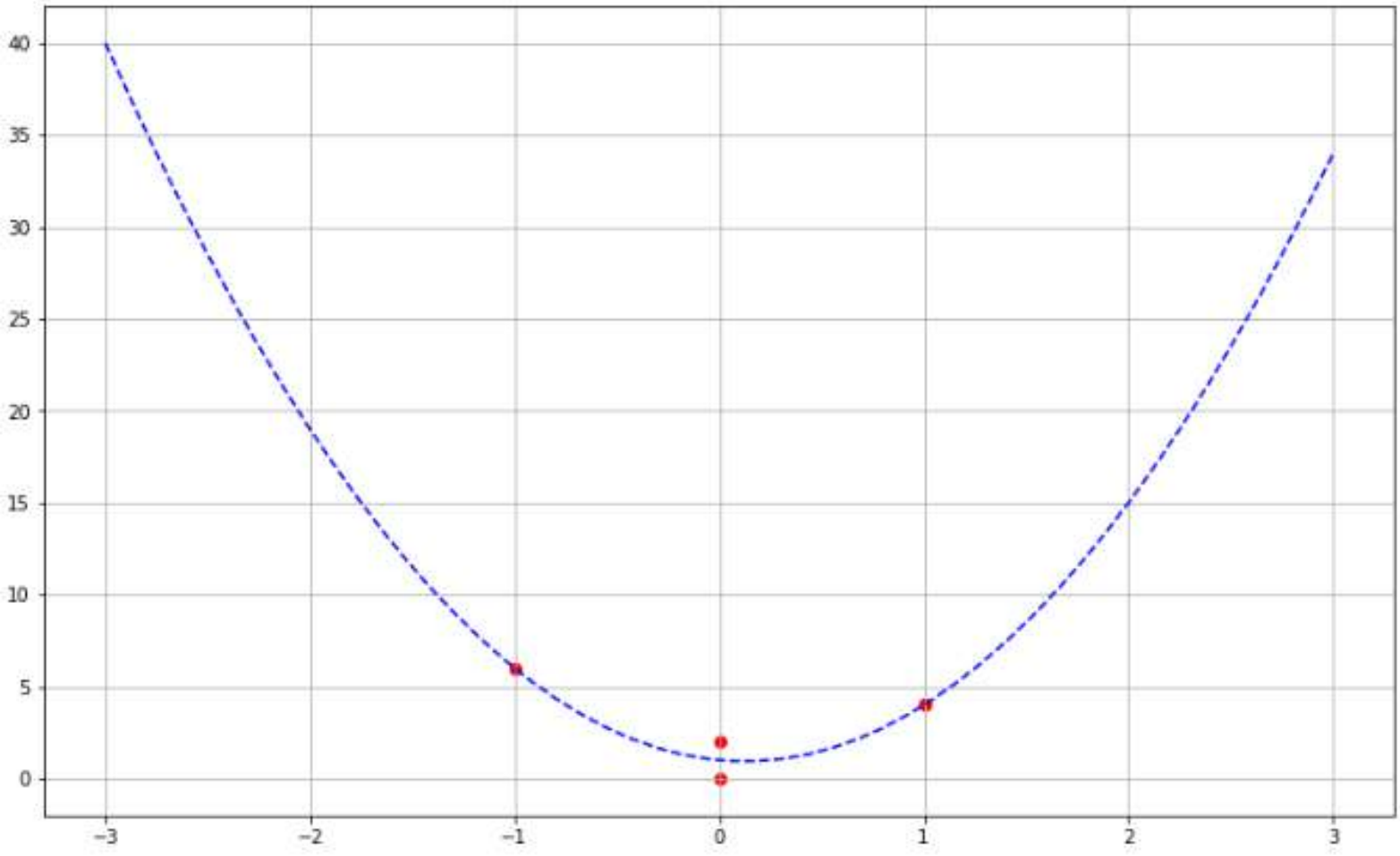


Картинка чуть ниже!

3) Ридж-регрессия с L2-регуляризацией ($\lambda = 1$)

$$\|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \to \min$$

$$X^T X + \lambda E = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix}$$


40
35
30
25
20
15
10
5
0
−3
−2
−1
0
1
2
3

$$\begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 4 \end{pmatrix} \rightsquigarrow \beta = \begin{pmatrix} 28 \\ \frac{1}{28} \\ -\frac{45}{28} \end{pmatrix}$$

$$f(x) = \frac{97}{28} + \frac{1}{28}x - \frac{45}{28}x^2$$

№42. Задача: Bin classification

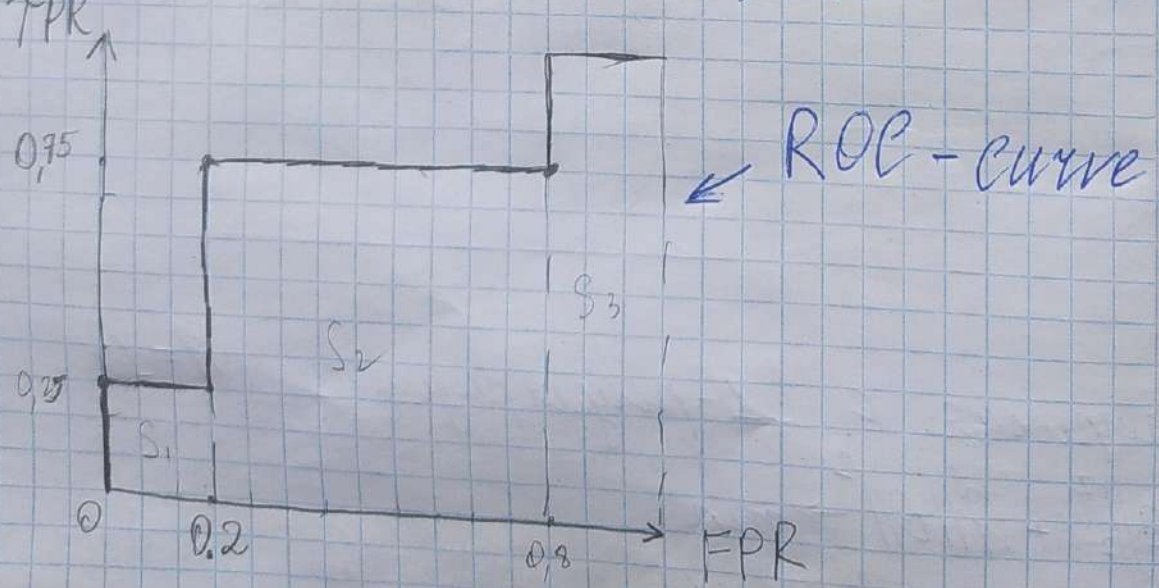
Предсказываем 1 для x.

i	1	2	3	4	5	6	7	8	9
$y^{(i)}$	0	0	0	0	0	1	1	1	1
$g(x^{(i)})$	0,75	0.15	0.11	0.23	0.09	0.1	0.66	0.82	0.5

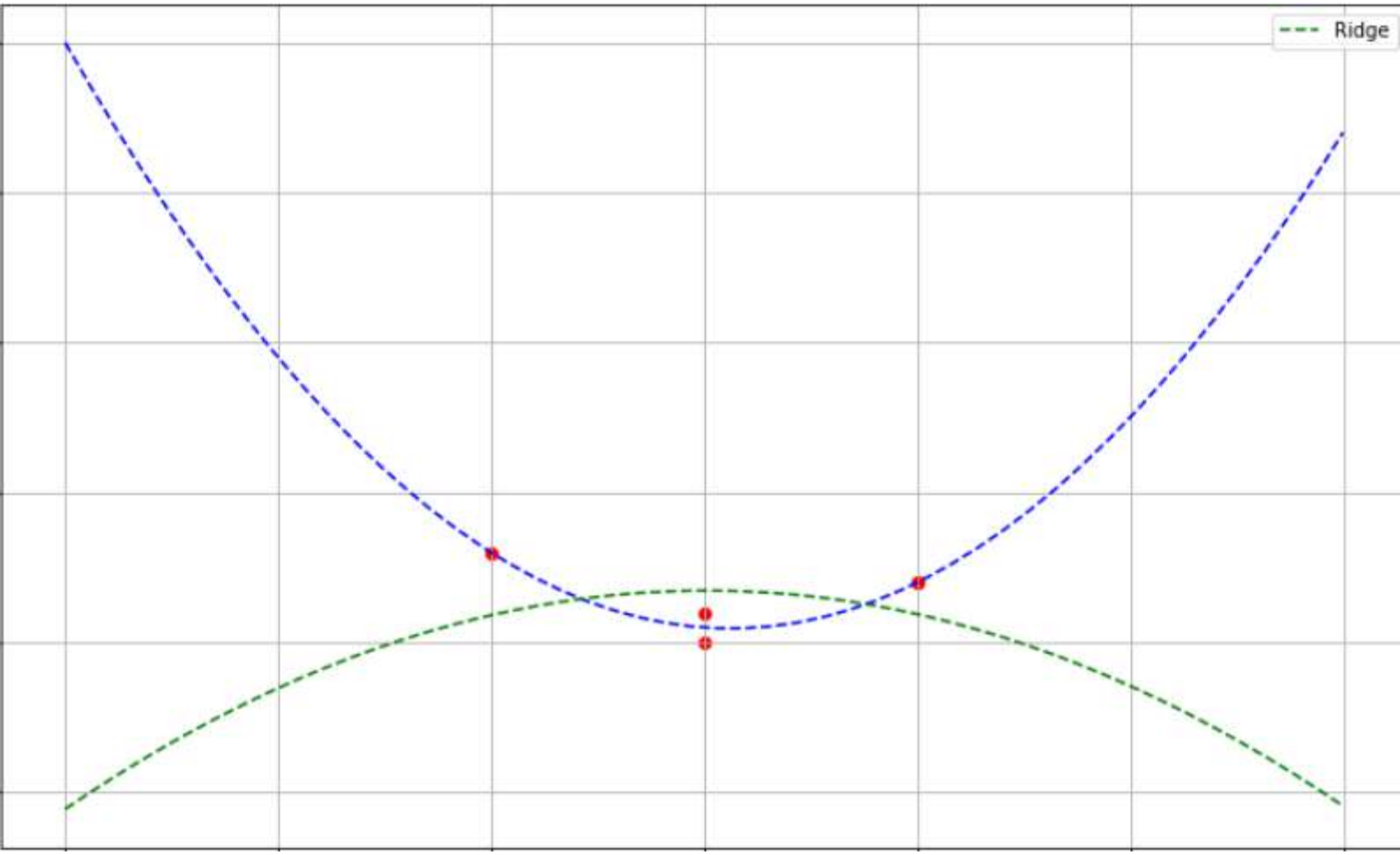
$g(x)$ - оценка апостериорной вероятности

FPR = [0, 0, 0.2, 0.2, 0.8, 0.8, 1]

TPR = [0, 0.25, 0.25, 0.75, 0.75, 1, 1]



$$AUC = S_1 + S_2 + S_3 = 0,2 \cdot 0,25 + 0,6 \cdot 0,75 + 0,2 = 0,7$$


Ridge

Для классификатора $f(x) = I(g(x) \geq 0.5)$ выписать матрицу рассогласования и найти необходимые метрики качества.

Confusion Matrix:

	$y=1$	$y=0$
$\hat{y}=1$	TP	FP
$\hat{y}=0$	FN	TN

$$Conf = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{7}{9}$$

$$error = 1 - accuracy = 2/9$$

$$FPR = \frac{FP}{FP+TN} = \frac{1}{5}$$

$$TPR = recall = \frac{TP}{TP+FN} = \frac{3}{4}$$

$$FNR = \frac{FN}{FN+TP} = \frac{1}{1+3} = \frac{1}{4}$$

$$PPV = \frac{TP}{TP+FP} = \frac{3}{4} = precision$$

$$TNR = \frac{TN}{TN+FP} = \frac{4}{5}$$

$$F_1 = (1+1) \cdot \frac{\frac{3}{4} \cdot \frac{3}{4}}{1 \cdot \frac{3}{4} + \frac{3}{4}} = \frac{3}{4}$$