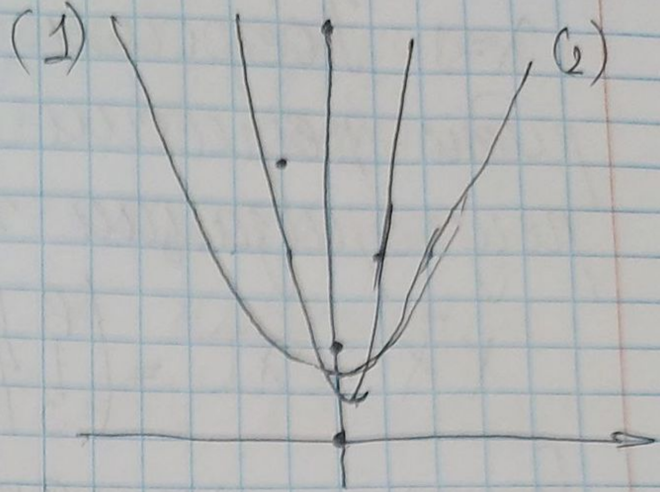
Дз №2

№3

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



$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \quad y = \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} \quad X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Решим систему $X^T X \beta = X^T y$

$$\begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

$$\begin{cases} \beta_0 = 1 \\ \beta_1 = -1 \\ \beta_2 = 4 \end{cases} \longrightarrow 1 - x + 4x^2 \quad (1)$$

$\lambda = 1$ Построим модель методом ридж-регрессии с параметром регуляризации λ.

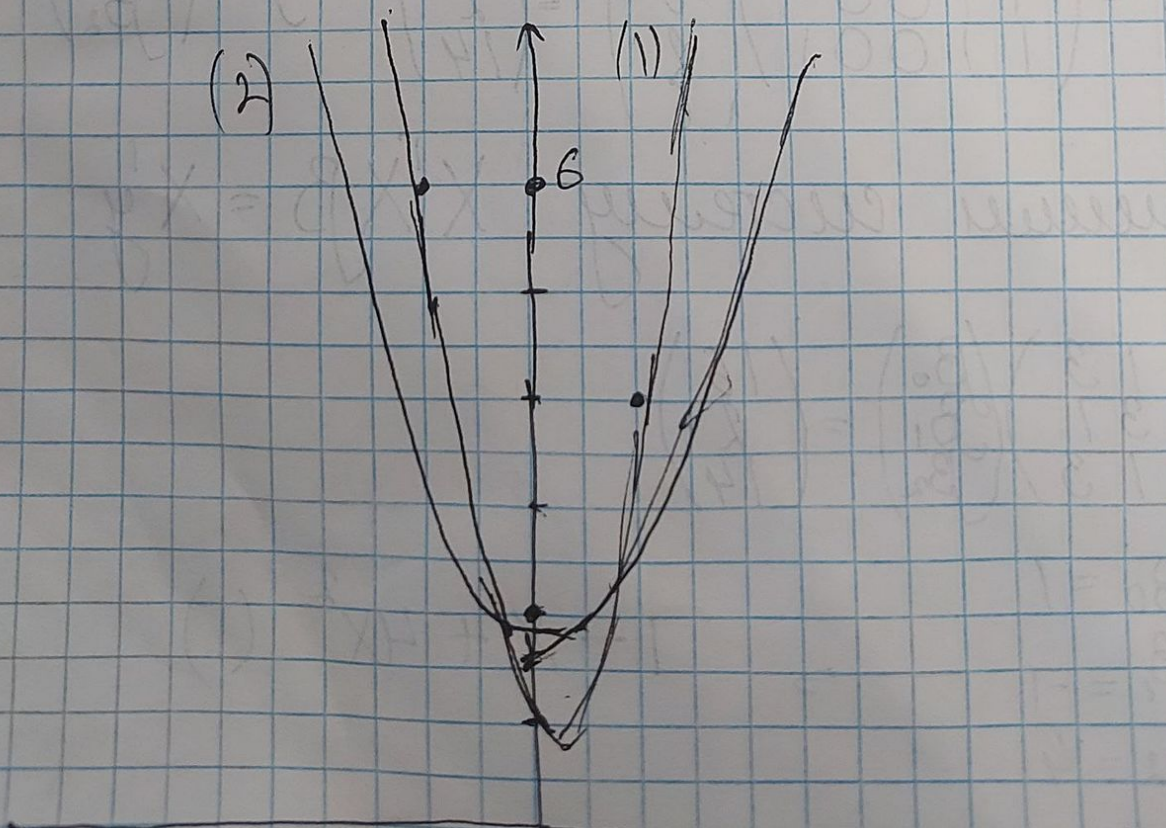
$$X^T X + \lambda I = \begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix}$$

Решаем регуляризованную систему нормальных урав-ний: $(X^T X + \lambda I)^{\wedge} \beta = X^T y$

$$\begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

$$\begin{cases} \beta_0 = 1{,}5 \\ \beta_1 = -0{,}5 \\ \beta_2 = 2{,}5 \end{cases} \longrightarrow$$ наим. модель

$1{,}5 - 0{,}5X + 2{,}5X^2$ (2)



№40 В задаче бинарной классификации, зная сколько в выборке положительных и отрицательных представителей и 2 показателя из TPR, TNR, PPV, NPV определить оставш. два.

∃ знаем TPR и TNR, а также $P = TP + FN$ и $N = TN + FP$, т.е находим

	$f(x) = 0$	$f(x) = 1$
$y = 0$	TN	FP
$y = 1$	FN	TP

$$PPV = \frac{TP}{TP + FP}, \quad NPV = \frac{TN}{TN + FN}$$

Выразим

$$TP = TPR(FN + TP) = TPR \cdot P$$

$$FN = FNR(FN + TP) = (1 - TPR)P$$

$$FP = FPR(TN + FP) = (1 - TNR)P$$

$$TN = TNR(TN + FP) = TNR \cdot N$$

Находим неизвестные показатели

$$PPV = \frac{TPR \cdot P}{TPR \cdot P + (1 - TNR) N}$$

$$NPV = \frac{TNR \cdot N}{TNR \cdot N + (1 - TPR) P}$$

$\Rightarrow$ можно вычислить PPV и NPV зная TPR и TNR, тоже можно доказать и для других пар показателей.

<u>№41</u> Не верно. PPV и TPR зависит только от положительных примеров, а TNR и NPV - только от отрицательных. Изменение FN или FP может повлиять на TNR и NPV не изменеяя PPV и TPR

2) Не верно, по тому же, что и 1) п.

3) Не верно RCC - кривая, отражает зависимость TPR от FPR. Precision от Recall.

ROC - кривая, показывает как классификатор различает положительные и отрицательные примеры.

№42 Посчитаем TPR и FPR

	Порог	TP	FP	FN	TN	TPR	FPR
	1	0	0	4	5	0	0
1 - 1,8 - 0	0,82	~~0~~ 1	0	3	5	0,25	0
2 - 1,7 - 0	0,75	1	1	3	4	0,25	0,2
3 - 1,6 - 0	0,66	2	1	2	4	0,5	0,2
4 - 1,5 - 0	0,5	3	1	1	4	0,75	0,2
5 - 1,4 - 0	0,23	3	2	1	3	0,75	0,4