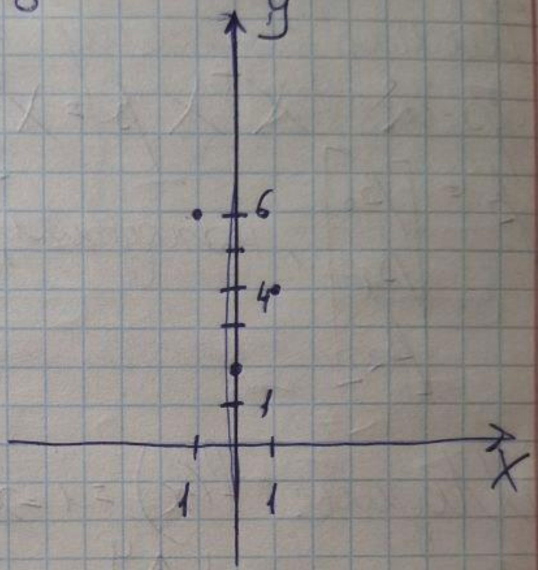
Кузьмин Александр, ПМ2

№3.

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



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

с помощью МНК найти $\beta_0, \beta_1, \beta_1$

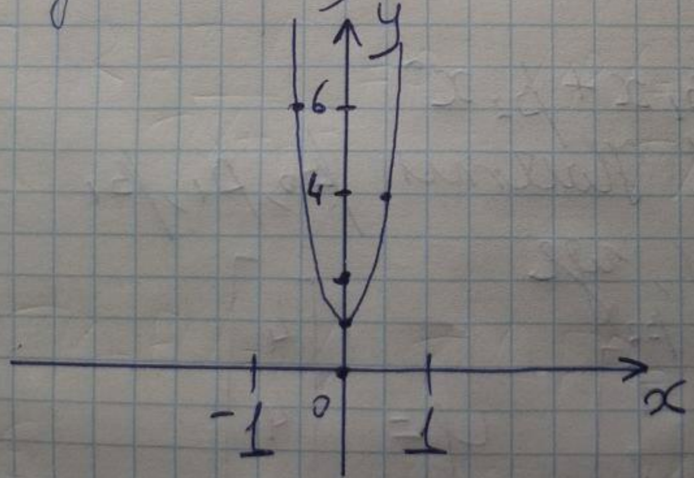
$X^T X \beta = X^T y$, где

$$X = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1 & \beta_2 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \end{array} \qquad y = \begin{bmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{bmatrix} \quad X^T y = \begin{bmatrix} 16 \\ 2 \\ 14 \end{bmatrix}$$

Решим систему нормальных ур-ий $X^T X \beta = X^T y$, где $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$, находим, что: $\beta = \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$

Модель: $f(x) = 1 - x + 4x^2$



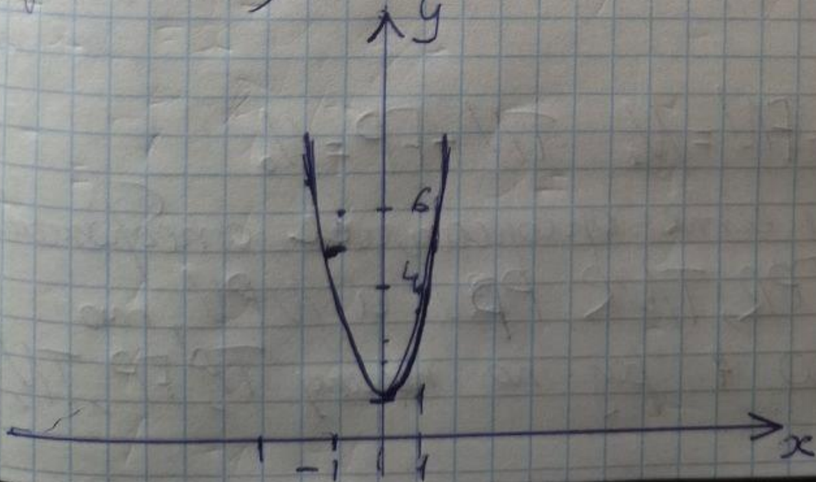
3) Построить модель того же вида методом ридж-регрессии с параметром регуляризации $\lambda = 1$; построить график этой ф-ии

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

Решаем регулярную систему:

$$(X^T X + \lambda \underbrace{I}_{\text{единичная матрица}})\beta = X^T y \; ; \; \beta = \begin{bmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix}$$

Модель: $f(x) = \frac{3}{2} - \frac{1}{2}x + \frac{5}{2}x^2$



N40

В задаче бин. классификации

известно: $N_1 + N_0 = N$, где N_1, N_0 – кол-во представителей одного класса

Два из 4-ёх показателя: TPR, TNR, PRV, NPV

У-тв: по ∀ двум можно посчитать остальные две.

Док-во: (в предположении, $TP \neq 0$, $TN \neq 0$)

1) $TPR = \frac{TP}{TP+FN} \neq 0$ $\quad PPV = \frac{TP}{TP+FP} \neq 0$

$TNR = \frac{TN}{TN+FP} \neq 0$ $\quad NPV = \frac{TN}{TN+FN} \neq 0$

2) $TP + FN = N_1$ $\quad TN + FP = N_0$

Если есть однозначное отображение м/у TPR, TNR, PPV и NPV, то можно 4 раз. по поиску TP, FP, TN, FN где знаем TPR, TNR, PRV и NPV

$$
\begin{cases}
N_1 \cdot TPR - TP = 0 \\
N_0 - TNP - TN = 0 \\
[TP + N_0 - TN]\, PPV - TP = 0 \\
-[TN + N_1 - TP]\, NPV - TN = 0 \\
TP + FN = N_1 \\
TN + FP = N_0
\end{cases}
$$

TP	TN	FP	FN	N_0	N_1	
$\boxed{1}$	0	0	0	0	$-TPR$	0
0	1	0	0	$-TNR$	0	0
$PPV-1$	$-PPV$	0	0	PPV	0	0
$-NPV$	$NPV-1$	0	0	0	NPV	0
1	0	$\boxed{1}$	0	0	-1	0
0	1	0	$\boxed{1}$	-1	0	0

TP	TN	FP	FN	N_0	N_1	
1	0	0	0	0	-TPR	0
0	1	0	0	-TNR	0	0
PN-1	-PN	0	0	PPV	0	0
-NPV	PN-1	0	0	0	NPV	0
1	0	$\boxed{1}$	0	0	-1	0
0	1	0	$\boxed{-1}$	-1	0	0

№41

Пусть расм. задача бинарной классификации

Верно ли, что:

1) если у 2-х классиф-ов на одной и той же выборке совпадает PPV и совпадают TPR, то будут совпадать TNR и NPV

Верное

$$PPV = \frac{TP}{TP+FP} = TPR = \frac{TP}{TP+FN}$$

$$\Rightarrow TP+FN = TP+FP \Rightarrow FN = FP$$

$$TNR = \frac{TN}{TN+FP} = NPV = \frac{TN}{TN+FN}$$

FN и FP равны из PRV = TPR

2) Верное

$$TNR = NPV \Rightarrow PPV = TPR$$

$$TNR = \frac{TN}{TN+FP} = NPV = \frac{TN}{TN+FN}$$

$\Rightarrow FP = FN$ (или $TN \neq 0$)

$$PPV = \frac{TP}{TP+FP} = TPR = \frac{TP}{TP+FN}$$

$\Rightarrow FP = FN$

3) Неверное

Совпадение ROC кривых влечёт совпадение PRECision-Recall кривых и наоборот.

ROC строится по точкам (FPR, TPR).

Если ROC совпадают $\not\to$ совпадают PR, т.к. используют разные метрики.

N42.

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{FP+TN}$$

$$F(x) = I(g(x) \geq 0{,}5)$$

$g(x) \geq 0{,}5$

$0{,}82\ (y=1) \to TP$

$0{,}76\ (y=0) \to FP$

$0{,}66\ (y=1) \to TP$

$0{,}5\ (y=1) \to TP$

actual \ predicted		
1	3	1
0	1	4

$g(x) < 0{,}5$:

$0{,}23\ (y=0) \to TN$

$0{,}15\ (y=0) \to TN$

$0{,}11\ (y=0) \to TN$

$0{,}10\ (y=1) \to FN$

$0{,}09\ (y=0) \to TN$

$TP = 3 \quad FN = 1$

$TP = 1 \quad TN = 4$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 0{,}778$$

$error = 1 - accuracy = 0{,}222$

$$F_1 = 2\,\frac{PPV \cdot TPR}{PPV+TPR} = 0{,}75$$

ROC-кривая $TPR = \frac{TP}{TP+FN}$

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

Упорядочим:

$g(x)$	y	TPR	FPR
0,82	1	0,25	0
0,75	0	0,25	0,2
0,66	1	0,5	0,2
0,5	1	0,75	0,2
0,23	0	0,75	0,4
0,15	0	0,75	0,6
0,11	0	0,75	0,8
0,10	1	1	0,8
0,09	0	1	1

$$TPR = \frac{TP}{TP+FN} = \frac{TP}{P}$$

$$FPR = \frac{FP}{FP+TN} = \frac{FP}{N}$$

Порог 0,82:

Все $g(x) \geq 0,82$ считаем как $\underline{1}$

Только один объект $(0,82, y=1)$

классифиц. как $\underline{1}$

$$TPR = \frac{1}{4} = 0,25$$

$$FPR = \frac{0}{5} = 0$$

Порог 0,75:

$g(x) \geq 0,75$

Два объекта $(0,82, y=1)$

$(0,75, y=0)$

$$TPR = \frac{1}{4} = 0,25$$

$$FPR = \frac{1}{5} = 0,2$$

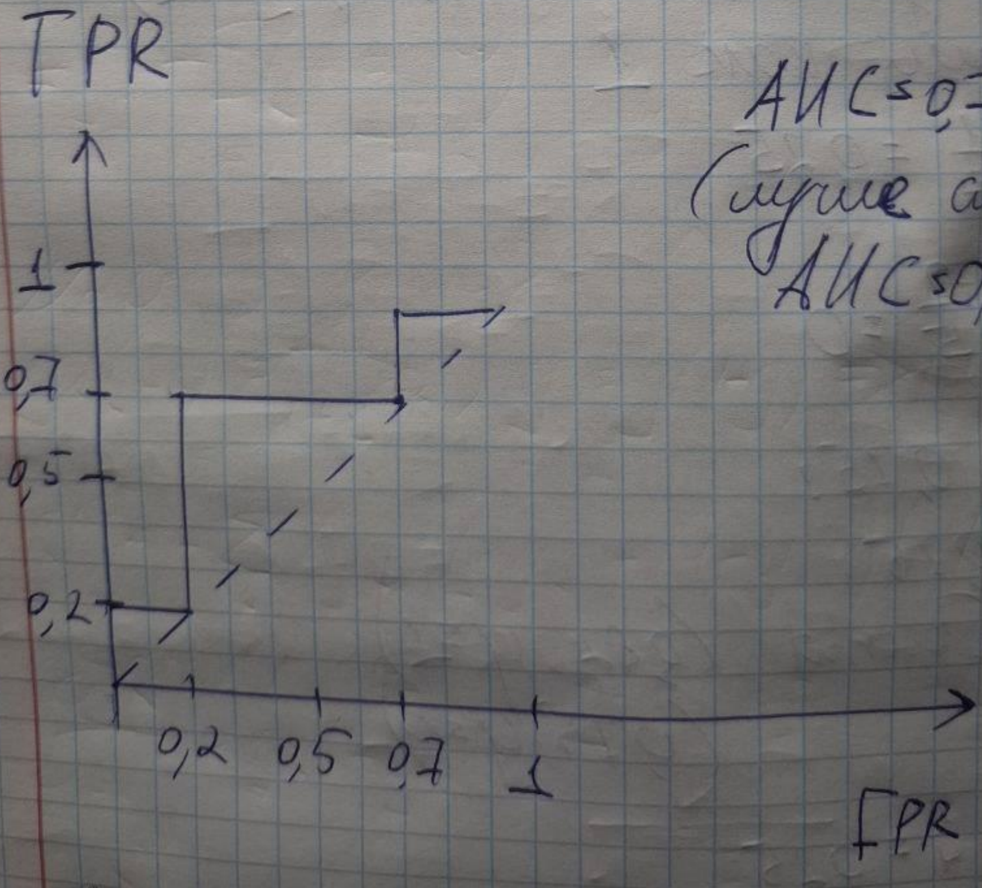
Порог 0,66:

$g(x) \geq 0,66$

Три объекта: $(0,62,\ y=1)$, $(0,75,\ y=0)$, $(0,66,\ y=1)$

$TPR = \frac{2}{4} = 0,5$

$FPR = \frac{1}{5} = 0,2$



AUC = 0,7
(лучше случайно
AUC = 0,5)