

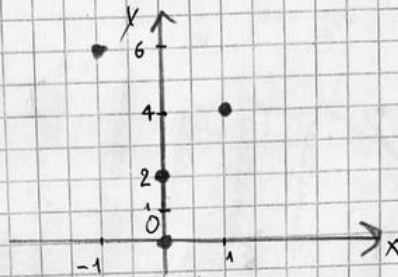
Крошечко Катя 382151ПМоп3

Линейная регрессия

③ Дана обучающая выборка

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

1) Изобразить точки:



2) Методом наименьших квадратов построить модель вида: $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, построить график этой ф-и

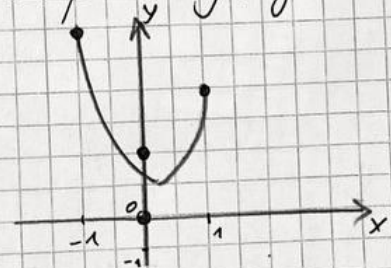
$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{bmatrix}$$

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

Решим сис-лу нормальных ур-й $X^T X \beta = X^T y$, где:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \text{ находим, что: } \beta = \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$$

Модель: $f(x) = 1 - x + 4x^2$

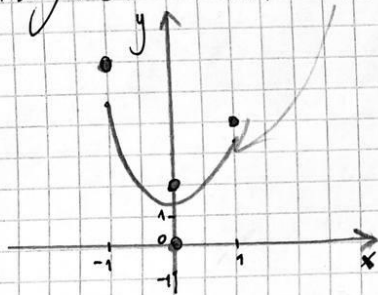


3) Построить модель того же вида методом ridge-регрессии с параметром регуляризации $\lambda = 1$; построить график этой ф-ии

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

решаем регуляризованную сис-му $(X^T X + \lambda I)\beta = X^T y : \beta = \begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix}$

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



I - единичная матрица

40] * задача бинарной классификации.

Док-ть, что если известно сколько в выборке представителей каждого из двух классов, то по 4-м показателям из списка TPR, TNR, PPV, NPV опр-е остальные два.

TN - true negative (истинно отрицат.)

TP - true positive (истинно полож.)

FP - false positive (ложно полож.)

FN - false negative (ложно отриц. пример)

промахи (ошибки 1-го рода, α -ошибки)

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

$$\text{error} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = 1 - \text{error}$$

$$\alpha = \text{FPR} = \frac{FP}{TN + FP} - \text{false positive rate}$$

$$\beta = \text{FNR} = \frac{FN}{FN + TP} - \text{false negative rate}$$

$$\text{precision} = \frac{TP}{FP + TP} - \text{точность}$$

$$\text{recall} = \text{TPR} = \frac{TP}{FN + TP} = 1 - \text{FNR}$$

$$\text{TNR} = \frac{TN}{TN + FP} = 1 - \text{FPR} - \text{specificity (специфичность)}$$

$$\text{TPR} = \frac{TP}{TP + FN} = 1 - \text{FNR}$$

$$\text{TNR} = \frac{TN}{TN + FP} = 1 - \text{FPR}$$

$$\text{PPV} = \frac{TP}{TP + FP}$$

$$\text{NPV} = \frac{TN}{TN + FN}$$

Два класса объектов известно
Positive class

• Общее кол-во объектов $N_1 = TP + FN$
 Negative class

• Общее кол-во объектов $N_2 = TN + FP$

$$\begin{cases} TPR = \frac{TP}{TP + FN} = \frac{TP}{N_1} \\ TNR = \frac{TN}{TN + FP} = \frac{TN}{N_2} \end{cases} \Rightarrow \begin{cases} TP = TPR \cdot N_1 \\ TN = TNR \cdot N_2 \end{cases}$$

$$\begin{cases} PPV = \frac{TP}{TP + FP} \\ NPV = \frac{TN}{TN + FN} \end{cases}$$

N_1, N_2 - given

$$\begin{cases} TP = TPR \cdot N_1 \\ TN = TNR \cdot N_2 \\ FP = \frac{TP \cdot (1 - PPV)}{PPV} \\ FN = \frac{TN \cdot (1 - NPV)}{NPV} \end{cases} \Rightarrow \begin{cases} TPR = \frac{TP}{N_1} \\ TNR = \frac{TN}{N_2} \\ PPV = \frac{TP}{TP + FP} \\ NPV = \frac{TN}{TN + FN} \end{cases}$$

Given:

1) TPR, TNR - trial cases.

$$TPR = \frac{TP}{N_1} \quad N_1 = TP + FN$$

$$TNR = \frac{TN}{N_2} \quad N_2 = TN + FP$$

Получив TP сразу найдем FN

Получив TN сразу найдем FP

TP, FN, TN, FP

2) PPV, NPV

$$N_1 = TP + \frac{FN(1 - NPV)}{NPV}$$

$$N_2 = TN + \frac{TP(1 - PPV)}{PPV}$$

Поиск TN, TP

Получив TN, TP из N_1, N_2 получим все значения

TP, FN, TN, FP \Rightarrow можно получить TPR, TNR

(41)] * задача димарной квалир-ии.
Верно м, что:

1) если у 2-х квалир-ов на одной и той же выборке совпадают 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 PPV = TPR

2) Верное

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

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

$$\Rightarrow FP = FN \text{ (если } TN \neq 0)$$

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

$\rightarrow FP = FN$

3) Неверное

Совпадение ROC кривых не влечет совпадение Precision-Recall кривых и наоборот

ROC строится по точкам (FPR, TPR)

Если ROC совпадает \rightarrow совпадает PR,

т.к. используют разные матрицы

42 $TPR = \frac{TP}{TP+FN} = FPR = \frac{FP}{FP+TN}$

$$f(x) = I(g(x) \geq 0,5)$$

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

$$0,82 (y=1) \rightarrow TP$$

$$0,76 (y=0) \rightarrow FP$$

$$0,66 (y=1) \rightarrow TP$$

$$0,5 (y=1) \rightarrow TP$$

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

$$0,23 (y=0) \rightarrow TN$$

$$0,15 (y=0) \rightarrow TN$$

$$0,11 (y=0) \rightarrow TN$$

$$0,10 (y=1) \rightarrow FN$$

$$0,09 (y=0) \rightarrow TN$$

$$FPR = \frac{FP}{FP+TN} = 0,2$$

$$FNR = \frac{FN}{TP+FN} = 0,25$$

$$TNR = \frac{TN}{TN+FP} = 0,8$$

$$TPR = \frac{TP}{TP+FN} = 0,75$$

$$PPV = \frac{TP}{TP+FP} = 0,75$$

	predicted		
	1	3	1
actual	0	1	4

$$TP = 3 \quad FN = 1$$

$$TP = 1 \quad TN = 4$$

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

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

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

Рос-кривая

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

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

Упорядочим

$g(x)$	y	TPR	FPR
0,82	1	0,25	0
0,78	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 1

Только один объект (0,82, $y=1$) классиф. как 1

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

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

Порог 0,75:

$$g(x) \geq 0,75$$

Два объекта $(0,82, y=1)$
 $(0,75, y=0)$

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

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

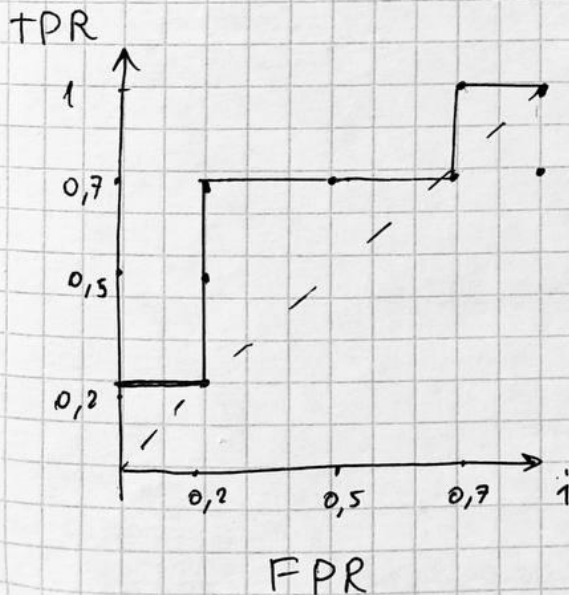
Порог 0,66:

$$g(x) \geq 0,66$$

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

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

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



$$0,2 \cdot 0,25 + 0,75 \cdot 0,6 + 1 \cdot 0,2 = 0,7$$

$$AUC = 0,7$$

(лучше случайного)
 $AUC = 0,5$