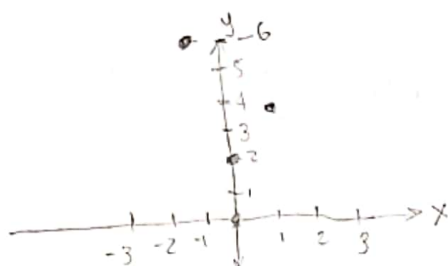


N3(1): Дана обучающая выборка

X	1	1	0	0	-1
Y	4	4	0	2	6

Влас Каролина
331903-1 ПИИ

1) Упорядочить точки:



N40: Пусть рассматривается задача бинарной классификации. Доказать, что если известно сколько в выборке представителей каждого из двух классов, то по любым двум показателям из списка TPR, TNR, PPV, NPV определяются остальные два.

X - представителей 1ого класса

Y - представителей 2ого класса

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

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

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

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

$$X+Y = TP+TN+FP+FN$$

$$TP+FN=X \quad FP+TN=Y \Rightarrow TPR = \frac{TP}{X} \quad TNR = \frac{TN}{Y}$$

1)] даны $TPR = k$ и $TNR = l$:

$$TP = kX \quad TN = lY$$

$$PPV = \frac{kX}{kX+FP}$$

$$NPV = \frac{lY}{lY+FN}$$

$$FN = X - TP = X - kX$$

$$FP = Y - TN = Y - lY$$

$$PPV = \frac{kX}{kX + (Y - lY)}$$

$$NPV = \frac{lY}{lY + (X - kX)}$$

2)] $TPR = k$ и $NPV = b$

$$TP = kX \quad b(TN+FN) = TN \rightarrow bTN + bFN = TN$$

$$TN = -\frac{bFN}{b-1} = -\frac{b(X-TP)}{b-1} = -\frac{b(X-kX)}{b-1}$$

$$PPV = \frac{kX}{kX+FP} = \frac{kX}{kX+(Y-TN)}$$

$$PPV = \frac{kX}{kX + \left(Y + \frac{b(X-kX)}{b-1} \right)}$$

$$TNR = \frac{-\frac{b(X-kX)}{b-1}}{(b-1)Y}$$

Для TNR и PPV все аналитично

$$3)] TPR = K \text{ и } PPV = \alpha$$

$$TP = Kx$$

$$FN = \frac{TP}{K} - TP \quad FP = \frac{TP}{\alpha} - TP$$

$$TNR = \frac{y - FP}{FP + (y - FP)} = \frac{y - \frac{TP}{\alpha} + TP}{\frac{TP}{\alpha} - TP + y - \frac{TP}{\alpha} + TP} = \frac{y - (\frac{Kx}{\alpha} - Kx)}{\frac{Kx}{\alpha} - Kx + y - \frac{Kx}{\alpha} + Kx} =$$

$$= \frac{y(\frac{Kx}{\alpha} - Kx)}{y}$$

$$NPV = \frac{y - FP}{y - FP + \frac{TP}{K} - TP} = \frac{y - \frac{TP}{\alpha} + TP}{y - \frac{TP}{\alpha} + \frac{TP}{K}} = \frac{y - \frac{Kx}{\alpha} + Kx}{y - \frac{Kx}{\alpha} + \frac{Kx}{K}} \quad \checkmark \text{ гораздо}$$

Далее TNR и NPV аннотировано

$$4)] PPV = \alpha \text{ и } NPV = b$$

$$a(TP + FP) = TP$$

$$aTP + aFP = TP$$

$$FP = \frac{TP - aTP}{a} =$$

$$= \frac{(x - FN)(1 - a)}{a} = \frac{(x - \frac{(y - FP)(1 - b)}{b})(1 - a)}{a} =$$

$$= \frac{(1 - a)x}{a} - \frac{(1 - a)(1 - b)(y - FP)}{ab}$$

$$abFP = (1 - a)x - (1 - a)(1 - b)(y - FP)$$

$$abFP = (1 - a)x - (1 - b - a + ab)(y - FP)$$

$$abFP = (1 - a)x - (y - by - ay + aby - FP + bFP + aFP - abFP)$$

$$abFP = x - ax - y + by + ay - aby + FP - bFP - aFP + abFP$$

$$FP(b + a - 1) = x - ax - y + by + ay - aby$$

$$FP = \frac{x - ax - y + by + ay - aby}{b + a - 1}$$

Отсюда находим FN

$$TP = x - FN \text{ и } TN = y - FP$$

✓ гораздо

Находим TPR и TNR по соответствующим формулам

N41) Пусть рассматривается задача бинарной классификации. Верно ли что:

- 1) Если у 2х классификаторов на одной и той же выборке совпадают PPV и TPR, то будут совпадать TNR и NPV

$$TPR = \frac{TP}{TP+FN} \Rightarrow TP \text{ для обоих классификаторов определяется единственным образом}$$

X . т.е. определ. единств. образом

$$PPV = \frac{TP}{TP+FP} \Rightarrow FP \text{ тоже определ. ед. образом. (TP - задано)}$$

$$FN = X - TP \quad TN = Y - FP \Rightarrow TNR \text{ и } NPV \text{ в обоих случаях будут одинаковы.}$$

- 2) TNR и NPV совпадают.

$$TNR = \frac{TN}{FP+TN} \Rightarrow TN \text{ определ. единств. образом}$$

$$NPV = \frac{TN}{TN+FN} \Rightarrow FN \text{ определ. ед. образом. (TN - задано)}$$

$$TP = X - FN \quad ; \quad FP = Y - TN \Rightarrow PPV \text{ и } TPR \text{ в обоих случаях будут одинаковыми.}$$

- 3) ROC кривых (совпадают)

ROC - кривая - график зависимости величин TPR и FPR

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

Precision - Recall - кривая

$$Recall = \frac{TP}{TP+FN} \quad Precision = \frac{TP}{TP+FP}$$

Как можно заметить, TPR и Recall вычисляются одинаково
 $\Rightarrow TP$ и FN характеристики определяются однозначно.

$$FP + TN = Y \Rightarrow FP \text{ тоже определяется однозначно} \Rightarrow$$

Совпадение ROC кривых влечет совпадение Precision-Recall кривых
 Обратное утверждение доказывается аналогично.

- N42) Пусть в задаче классификации на 2 класса $\{0, 1\}$ некоторый классификатор определяет следующие оценки $g(x)$ альтернативной вероятности принадлежности объектов x к классу 1:

i	1	2	3	4	5	6	7	8	9
$g^{(i)}$	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	0,50
$\geq 0,5$	1	1	1	1	0	0	0	0	0

Постройте ROC-кривую.

Выпишите AUC.

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

Выпишите матрицу

ассоциаций и найдите

FPR, FNR, TNR, TPR, PPV, accuracy, error, F1

	y=1	y=0
$\hat{y}=1$	3 (TP)	1 (FP)
$\hat{y}=0$	1 (FN)	4 (TN)

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

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

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

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

$$AUC = \frac{1}{5} \cdot \frac{1}{4} + \frac{1}{5} \cdot 3 \cdot \frac{1}{4} \cdot 3 +$$

$$+ \frac{1}{5} \cdot 4 \cdot \frac{1}{4} = \frac{1}{20} + \frac{9}{20} + \frac{4}{20} = \frac{14}{20} = 0,7$$

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

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

$$error = \frac{FP+FN}{TP+TN+FP+FN} = \frac{2}{9}$$

$$F1 = 2 \cdot \frac{recall \times precision}{recall + precision} = 2 \cdot \frac{\frac{3}{4} \times \frac{3}{4}}{\frac{3}{4} + \frac{3}{4}} = \frac{2 \cdot 9 \cdot 1}{16 \cdot 6_2} = \frac{3}{4}$$

