

№ 15

Шкелева К.

x_1	0	0	1	1	0	0	1	1	1	0
x_2	0	1	0	1	1	1	1	1	1	1
y	0	0	0	0	0	1	1	1	1	1

Оценим априорные вероятности:

$$Pr^1(Y=0) = \frac{1}{2} \quad Pr^1(Y=1) = \frac{1}{2}$$

Оценим условные вероятности:

$$Pr^1(x_1=0/Y=0) = \frac{3}{5} \quad Pr^1(x_2=0/Y=0) = \frac{2}{5}$$

$$Pr^1(x_1=1/Y=0) = \frac{2}{5} \quad Pr^1(x_2=1/Y=0) = \frac{3}{5}$$

$$Pr^1(x_1=0/Y=1) = \frac{2}{5} \quad Pr^1(x_2=0/Y=1) = 0$$

$$Pr^1(x_1=1/Y=1) = \frac{3}{5} \quad Pr^1(x_2=1/Y=1) = 1$$

Используя основное предположение наивного байесова классификатора, получаем

$$\begin{aligned} Pr(Y=0 | x_1=1, x_2=1) &= \frac{Pr(x_1=1/Y=0) \cdot Pr(x_2=1/Y=0) \cdot Pr(Y=0)}{Pr(x_1=1, x_2=1)} = \\ &= \frac{\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{\frac{3}{25}}{\frac{21}{50}} = \frac{3 \cdot 50^2}{21 \cdot 25} = \frac{2}{7} \end{aligned}$$

$$\begin{aligned} Pr(Y=1 | x_1=1, x_2=1) &= \frac{Pr(x_1=1/Y=1) \cdot Pr(x_2=1/Y=1) \cdot Pr(Y=1)}{Pr(x_1=1, x_2=1)} = \\ &= \frac{\frac{3}{5} \cdot 1 \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{\frac{3}{10}}{\frac{21}{50}} = \frac{3 \cdot 50^5}{21 \cdot 10} = \frac{5}{7} \end{aligned}$$

~ 40

Если известно, сколько в выборке представителей каждого из двух классов, то известны N и P .

Пусть известны TPR и TNR

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

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

$$FN = \frac{TP \cdot (1-TPR)}{TPR} = \frac{TPR \cdot P \cdot (1-TPR)}{TPR} = P \cdot (1-TPR)$$

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

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} \rightarrow TN = TNR \cdot N$$

$$FP = \frac{TN \cdot (1-TNR)}{TNR} = \frac{TNR \cdot N \cdot (1-TNR)}{TNR} = N \cdot (1-TNR)$$

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

← Получили 2 уравнения с двумя неизвестными PPV и NPV .

В общем случае получили 2 уравнения с 4 неизвестными. Если 2 из них известны, то остальные 2 можно найти.

~ 41

1) $TPR = PPV$

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

да, верно.

2) то же самое, что в п.1, только наоборот.

да, верно.

3) да, верно.

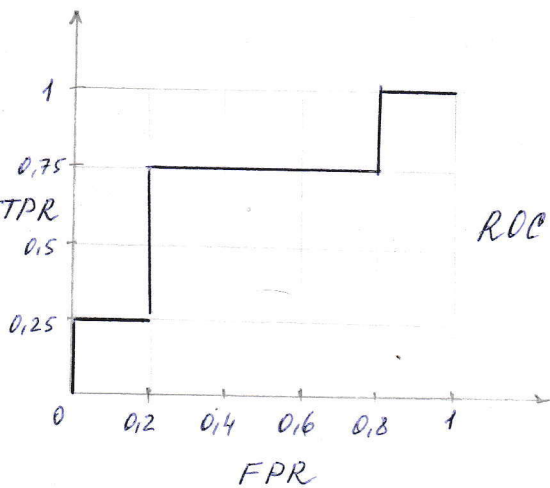
N 42

Шкунева К.

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	0,10	0,66	0,82	0,50

Отсортируем по $p(x^1) \leq p(x^2) \leq \dots \leq p(x^9)$

i	5	6	3	2	4	9	7	1	8
y_i	0	1	0	0	0	1	1	0	1
$g(x_i)$	0,09	0,10	0,11	0,15	0,23	0,50	0,66	0,75	0,82



ROC-кривая

AUC = площадь под ROC

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

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

	$y=1$	$y=0$
$g(x) \geq 0,5$	TP 3	FP 1
$g(x) < 0,5$	FN 1	TN 4

$$TP=3 \quad FP=1 \quad FN=1 \quad TN=4$$

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

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

$$TNR = 1 - FPR = 1 - \frac{1}{5} = \frac{4}{5}$$

$$TPR = 1 - FNR = 1 - \frac{1}{4} = \frac{3}{4}$$

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

$$\text{accuracy} = \frac{TP+TN}{TP+FP+FN+TN} = \frac{3+4}{3+1+1+4} = \frac{7}{9}$$

$$\text{error} = 1 - \text{accuracy} = 1 - \frac{7}{9} = \frac{2}{9}$$

$$F1 = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR} = \frac{2 \cdot \frac{3}{4} \cdot \frac{3}{4}}{\frac{3}{4} + \frac{3}{4}} = \frac{3}{4}$$