Шкилева К.

№4. $y \sim N(X\beta, \sigma^2 I)$
$\beta \sim N(0, \tau I)$

Пусть входные данные центрированы. Тогда апостериорное распределение для β имеет вид:

$$Pr(\beta|y, X) = \frac{1}{K} Pr(y|\beta, X) Pr(\beta)$$

где $K = K(y, X) = \int Pr(y|\beta, X) Pr(\beta) d\beta$ явно не зависит от β.

$$Pr(\beta|y, X) = \frac{1}{Z} \cdot \frac{1}{\sqrt{(2\pi)^p}\sigma^p} \cdot \exp\left(-\frac{(y-X\beta)^T(y-X\beta)}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{(2\pi)^p}\tau^p} \cdot \exp\left(-\frac{\beta^T\beta}{2\tau^2}\right) \quad (*)$$

Тогда $$\log(Pr(\beta|y, X)) = -C - \frac{(y-X\beta)^T(y-X\beta)}{2\sigma^2} - \frac{\beta^T\beta}{2\tau^2}$$

где C — члены, которые не зависят от β. Это выражение достигает максимума в $\hat\beta$:

$$\hat\beta = \left(X^TX + \frac{\sigma^2}{\tau^2} I\right)^{-1} X^T y$$

Если положим $\lambda = \frac{\sigma^2}{\tau^2}$, то увидим связь с гребневой регрессией.

Ясно, что $Pr(\beta|y, X)$ является распределением Гаусса и его среднее и мода совпадают. Покажем, что среднее значение равно $\hat\beta$.

Заметим, что из $(*)$ следует, что ковариация Σ равна

$$\Sigma^{-1} = \frac{1}{\sigma^2}\left(X^TX + \frac{\sigma^2}{\tau^2} I\right)$$

Отсюда следует $\hat\beta = \frac{1}{\sigma^2} \Sigma X^T y$. Приравнивая соответствующие члены в $(*)$, видим, что это должно быть среднее.

Шкилева К.

№ 15,

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:

$\hat{Pr}(Y=0) = \frac{1}{2}$ $\qquad \hat{Pr}(Y=1) = \frac{1}{2}$

Оценим условные вероятности:

$\hat{Pr}(X_1=0 / Y=0) = \frac{3}{5}$ $\qquad \hat{Pr}(X_2=0/Y=0) = \frac{2}{5}$

$\hat{Pr}(X_1=1 / Y=0) = \frac{2}{5}$ $\qquad \hat{Pr}(X_2=1/Y=0) = \frac{3}{5}$

$\hat{Pr}(X_1=0 / Y=1) = \frac{2}{5}$ $\qquad \hat{Pr}(X_2=0/Y=1) = 0$

$\hat{Pr}(X_1=1 / Y=1) = \frac{3}{5}$ $\qquad \hat{Pr}(X_2=1/Y=1) = 1$

Используя основное предположение наивного байесова классификатора, получаем

$$Pr(Y=0 / X_1=1, X_2=1) = \frac{Pr(X_1=1/Y=0)\cdot Pr(X_2=1/Y=0)\cdot Pr(Y=0)}{Pr(X_1=1, X_2=1)} =$$

$$= \frac{\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{1}{2}}{\frac{3}{25}+\frac{3}{10}} = \frac{\frac{3}{25}}{\frac{21}{50}} = \frac{3\cdot 50}{21\cdot 25} = \frac{2}{7}$$

$$Pr(Y=1 / X_1=1, X_2=1) = \frac{Pr(X_1=1/Y=1)\cdot Pr(X_2=1/Y=1)\cdot Pr(Y=1)}{Pr(X_1=1, X_2=1)} =$$

$$= \frac{\frac{3}{5}\cdot 1\cdot\frac{1}{2}}{\frac{3}{25}+\frac{3}{10}} = \frac{\frac{3}{10}}{\frac{21}{50}} = \frac{3\cdot 50}{21\cdot 10} = \frac{5}{7}$$

Шкилева К.

№ 40,

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

Пусть известны TPR и TNR

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

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

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

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

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

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

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

Получили 2 уравнения с двумя неизвестными PPV и NPV.

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

№ 41,

1) $TPR = PPV$

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

да, верно.

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

да, верно.

3) да, верно.

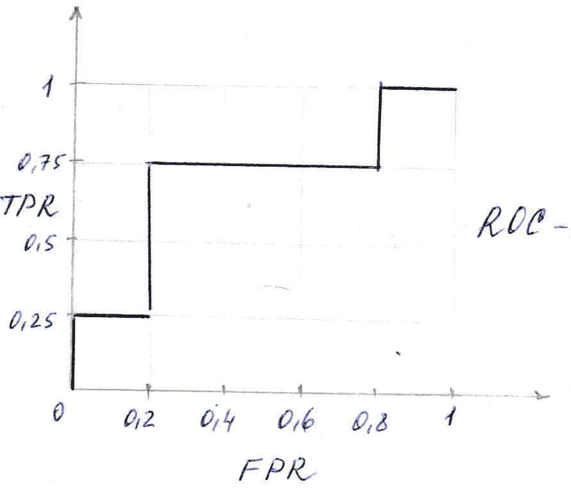
Шкилева К.

№42

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 $\rho(x^1) \le \rho(x^2) \le \ldots \le \rho(x^9)$

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

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

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

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

TP = 3   FP = 1   FN = 1   TN = 4

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

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

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

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

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

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

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

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