48 22 55 44 300

Materiały dla Wojciecha Nieminy z Mat. Stos.
od Piotra Pokarowskiego

Zad	Punktacja
1	3
2	3
3	2
4	2
5	3

liczba punktów	0 ÷ 5	5.5 – 6	6.5 – 8	8.5 – 10	10.5 – 12	12.5 – 13
ocena	2	3	3+	4	4+	5

Maksymalna rozdzielczość 0.5 punktu.

Zad 1. Wyprowadź wzór na drugą zmienną kanoniczną.

Rozw. X - macierz danych $n \times p$ (n obserw. p-wym.)
$T := \mathrm{var}(X)$, W - macierz wariancji wewnątrzgrupowych, B - macierz wariancji pomiędzygrupowych; $T = W + B$.
Zadanie polega na znalezieniu dwóch pierwszych kierunków kanonicznych: 1° $a_1^T B a_1 \to \max\limits_{a_1}$, $a_1^T W a_1 = 1$

2° $a_2^T B a_2 \to \max\limits_{a_2}$, $a_2^T W a_2 = 1$, $a_1^T W a_2 = 0$.

Mając a_2 wyznaczamy 2-gą zmienną kanoniczną $z_2 = X a_2$, (studenci mylą z_2 i a_2 - możemy pozostać na wyzn. a_2).

(i) W ma rozkład Choleskiego $U^T U$ (zakładamy, że $\exists W^{-1}$)
Zatem dla $i = 1, 2$: $a_i^T B a_i = a_i^T U^T (U^T)^{-1} B U^{-1} U a_i$ oraz $a_i^T U^T U a_i = 1$. Podstawiając $C := (U^T)^{-1} B U^{-1}$ oraz $u_i = U a_i$ mamy zadanie analizy składowych głównych:

(PCA) $u_i^T C u_i \to \max\limits_{u_i}$, $u_i^T u_i = 1$, $u_1^T u_2 = 0$.

(ii) Rozwiązujemy PCA. Niech $V \Lambda V^T$ będzie rozkładem spekt. C, gdzie $V = [v_1, \dots, v_p]$ ortogonalna oraz $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_1 \geq \dots \geq \lambda_p \geq 0$. Przyjmijmy, że $u = \sum\limits_j \alpha_j v_j$. Wtedy
$u^T C u = \left(\sum\limits_j \alpha_j v_j^T\right)\left(\sum\limits_k \alpha_k \lambda_k v_k\right) = \sum\limits_j \lambda_j \alpha_j^2 =:$ Prawa.

Ponieważ $\sum\limits_j \alpha_j^2 = u^T u = 1$, więc Prawa $\leq \lambda_1$. Maksimum jest realizowane przez v_1, natomiast wśród u ortogonalnych do v_1 maksimum jest realizowane przez v_2 oraz jest równe λ_2.

(iii) Sprawdzamy, że $a_i := U^{-1} v_i$ (bo $u_i = U a_i$) spełniają ograniczenia, np.: $a_1^T U^T U a_2 = v_1^T v_2 = 0$.

Zad 4. Wyprowadź wzór na optymalną regułę decyzyjną (względem "0-1" f-cji straty) w wielowymiarowym modelu normalnym: $N(m_1, \Sigma_1), \ldots, N(m_k, \Sigma_k)$ z rozkładem a-priori (q_i)

Rozw (i) Optymalna reguła decyzyjna (klasyfikacyjna) to

$$d(x) = \underset{1 \le j \le k}{\operatorname{argmax}}\, p(j|x) = \underset{j}{\operatorname{argmax}} \log\left[f(x|j)\, q_j\right].$$

(ii) Podstawiamy gęstość normalną za $f(x|j)$ i upraszczamy wyrażenie (rozkłady norm. są p-wymiarowe)

$$d(x) = \underset{j}{\operatorname{argmax}}\left[-\frac{p}{2}\log\Sigma_j - \frac{1}{2}(x-m_j)^T\Sigma_j^{-1}(x-m_j) - \log q_j\right]$$

Zad 5 Niech $\hat\beta$ będzie estymatorem najm. kwadr. wyznaczonym w modelu regresji liniowej wielu zmiennych z n-niezależnych obserwacji $(y_1, x_1), \ldots, (y_n, x_n)$. Niech ponadto (y, x) będzie $(n+1)$szą obserwacją niezależną od pozostałych. Oblicz:

(a) $\operatorname{var}(y_i - \hat\beta^T x_i)$ dla $i = 1, \ldots, n$;

(b) $\operatorname{var}(y - \hat\beta^T x)$

Rozw (a) Z wykładu: jeśli $H := X(X^TX)^{-1}X^T$, $\hat Y := HY$, X-macierz eksperymentu $\begin{bmatrix} 1 & x_1 \\ & \vdots \\ & x_n \end{bmatrix}$, to

$$\sigma^2 I = \operatorname{var} Y = \sigma^2 H + \sigma^2(I - H) = \operatorname{var}\hat Y + \operatorname{var}(Y - \hat Y).$$

Ponieważ $\hat\beta = (X^TX)^{-1}X^TY$, więc $\hat Y = X\hat\beta$.

Czyli $\operatorname{var}(y_i - \hat\beta^T x_i)$ to element (i,i) macierzy $\operatorname{var}(Y - \hat Y) = \sigma^2(I - H)$. Zatem $\operatorname{var}(y_i - \hat\beta^T x_i) = \sigma^2(1 - x_i^T(X^TX)^{-1}x_i)$.

(b) Wiemy z wykładu, że $\operatorname{var}\hat\beta = \sigma^2(X^TX)^{-1}$.

Zatem z niezależności $\operatorname{var}(y - \hat\beta^T x) = \operatorname{var} y + \operatorname{var}\hat\beta^T x$

$= \sigma^2 + x^T(\operatorname{var}\hat\beta)x = \sigma^2(1 + x^T(X^TX)^{-1}x)$.

Za zrobienie (a) - 1pkt, za (b) - 2pkty.

Zad 2 Niech X, Y będą zm. los. oraz f - funkcja X mierzalna. Udowodnij, że $cor(Y, E(Y|X))^2 \leq cor(Y, f(X))^2$.

Rozw (i) Niech $\hat{Y} \equiv E(Y|X)$ oraz $cor(X, Y) \equiv \rho(X, Y)$.

$$cov(Y, f) = E(Y - EY)(f - Ef) = EY(f - Ef) = E((f - Ef)\hat{Y})$$
$$= E(f - Ef)(\hat{Y} - E\hat{Y}) = cov(\hat{Y}, f)$$

(ii) W szczególności dla $f = \hat{Y}$ mamy $cov(Y, \hat{Y}) = var\hat{Y} =: \sigma^2_{\hat{Y}}$

(iii) Zatem $\rho(Y, f) = \dfrac{cov(Y, f)}{\sigma_Y \sigma_f} = \dfrac{\sigma^2_{\hat{Y}}}{\sigma_Y \sigma_{\hat{Y}}} \dfrac{cov(\hat{Y}, f)}{\sigma_{\hat{Y}} \sigma_f}$

$= \dfrac{cov(Y, \hat{Y})}{\sigma_Y \sigma_{\hat{Y}}} \dfrac{cov(\hat{Y}, f)}{\sigma_{\hat{Y}} \sigma_f}$. Po podniesieniu stronami do kwadratu dostajemy:

$\rho^2(Y, f) = \rho^2(Y, \hat{Y})\rho^2(\hat{Y}, f)$. Oczywiście $\rho^2(Y, \hat{Y}) \leq 1$, więc maksimum lewej mamy dla f t, że $\rho^2(\hat{Y}, f) = 1$, a więc również dla $f = \hat{Y}$.

Zad 3 Sformułuj model logistyczny i podaj algorytm dla wyznaczenia jego parametrów z zasady największej wiarygodności.

Rozw (i) $(Y_i, X_i)_1^n$ niezal., $Y_i \sim binom(1, p(x))$, gdzie

$$logit\, p(x) = \alpha + \beta^T x \iff p(x) = \frac{e^{\alpha + \beta^T x}}{1 + e^{\alpha + \beta^T x}}$$

(ii) Wiarygodność próby: $L(\alpha, \beta) := \prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i))^{1 - y_i}$

Log-wiarygodność $l(\alpha, \beta) := \sum_i y_i(\alpha + \beta^T x_i) + \sum_i (1 - y_i) - \sum_i \log \Delta_i$

gdzie $\Delta_i = 1 + e^{\alpha + \beta^T x_i}$. Parametry α, β znajdujemy numerycznie, np za pomocą alg. Newtona.