

Analiza Dużych Zbiorów Danych

Lista nr.2 - Wielokrotne testowanie

Dla $p = 1000$ wygeneruj ortonormalną macierz planu $X_{1000 \times 1000}$, tzn. taką macierz, że $X^T X = I_{p \times p}$. Następnie wygeneruj wektor współczynników regresji jako ciąg niezależnych zmiennych losowych z rozkładu

$$\beta_i \sim (1 - \epsilon)\delta_0 + \epsilon\phi(0, \tau^2) ,$$

gdzie δ_0 jest rozkładem skupionym w 0 a $\phi(0, \tau^2)$ jest gęstością rozkładu normalnego $N(0, \tau^2)$. Rozważ przypadki : $\epsilon \in \{0.01, 0.05, 0.1\}$ i $\tau \in \{1.5\sqrt{2 \log 1000}, 3\sqrt{2 \log 1000}\}$.

Dla każdego z tych przypadków wygeneruj wektor odpowiedzi

$$Y = X\beta + \epsilon ,$$

gdzie $\epsilon \sim N(0, I_{1000 \times 1000})$ i przeprowadź poniższe analizy.

1. Podaj wzór na estymator najmniejszych kwadratów β^{LS} dla wektora β i rozkład tego estymatora.
2. Zakładając, że wariancja błędu jest znana ($\sigma^2 = 1$), skonstruuj oba estymatory Jamesa-Steina dla wektora parametrów β (tzn. estymator ściągający do zera i do wspólnej średniej).
3. Zastosuj następujące procedury do ustalenia które zmienne są istotne
 - i) klasyfikator Bayesowski przy założeniu tej samej funkcji straty za błąd pierwszego i drugiego rodzaju
 - ii) procedurę Bonferoniego
 - iii) procedurę Benjaminiego-Hochberga.

Konstrukcje w ii) i iii) przeprowadź zakładając znajomość $\sigma = 1$.

Następnie dla każdego z punktów i), ii) i iii) wyznacz "ucięte" estymatory β

$$\hat{\beta}_i = \begin{cases} \beta^{LS} & \text{jeżeli odrzucono } H_{0i} : \mu_i = 0 \\ 0 & \text{w przeciwnym wypadku} \end{cases} .$$

- a) Estymatory z punktów 1-3 (6 estymatorów) porównaj pod kątem błędu kwadratowego

$$SE = ||\hat{\beta} - \beta||^2 .$$

- b) Procedury testowania z punktu 3 porównaj pod kątem sumy liczby błędów pierwszego i drugiego rodzaju.
- c) Dla każdej kombinacji ϵ i τ powtórz doświadczenie 1000 razy i porównaj analizowane procedury pod kątem $MSE = E(SE)$ i wartości oczekiwanej sumy liczby błędów pierwszego i drugiego rodzaju.

Malgorzata Bogdan