Duże bazy danych

Ridge regression

1. Generate orthonormal ($X'X = I$) matrix od dimension 1000×950. Consider the regression model

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

with $\epsilon \sim N(0, I_{n\times n})$ and the vector of regression coefficients $\beta_1 = \ldots = \beta_k = 3$ and $\beta_{k+1}, \ldots, \beta_{950} = 0$ with

   a) $k = 20$,

   b) $k = 100$,

   c) $k = 200$.

   For each of these cases

   i) Calculate the value of the tuning parameter λ for the ridge regression, so as to minimize the mean square error of the estimation of β.

   ii) Calculate the bias, the variance and the mean squared error of this optimal estimate.

   iii) Find the critical value and calculate the power of the statistical test based on the ridge estimator and controlling FWER at the level 0.1.

   iv) Generate 200 replicates of the above model and analyse the data using ridge regression and OLS. Compare empirical bias, variance, mse and the power of the test based on the ridge estimator with the theoretical values of these parameters, calculated above and with the corresponding parameters of OLS.

Malgorzata Bogdan