

# Advanced methods of statistical learning

## Shrinkage methods and knockoffs

1. Generate the design matrix  $X_{500 \times 450}$  such that its elements are iid random variables from  $N(0, \sigma = \frac{1}{\sqrt{n}})$ . Then generate the vector of the response variable according to the model

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

where  $\epsilon \sim 2N(0, I)$ ,  $\beta_i = 10$  for  $i \in \{1, \dots, k\}$  and  $\beta_i = 0$  for  $i \in \{k+1, \dots, 450\}$  and  $k \in \{5, 20, 50\}$ .

For 100 replications of the above experiments estimate the regression coefficients and/or identify important variables using

- i) least squares
- ii) ridge regression and LASSO with the tuning parameters selected by cross-validation
- iii) use knockoffs with ridge and LASSO to identify important variables while keeping FDR equal to 0.2.
- iv) adaptive LASSO I:  
first step: calculate weights using cross-validated LASSO (eliminate variables not selected by cross-validated LASSO)  
second step: use weighted cross-validated LASSO
- v) adaptive LASSO II: first step: calculate  $\hat{\beta}$  using cross-validated LASSO  
second step: estimate  $\sigma$

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-k}} ,$$

where  $RSS$  is from the cross-validated LASSO and  $k$  is the number of variables selected by cross-validated LASSO third step: calculate weights  $w_i = \frac{\hat{\sigma}}{|\hat{\beta}_i|}$

fourth step: use weighted LASSO with the tuning parameter  $\lambda = \hat{\sigma} \Phi^{-1} \left(1 - \frac{0.2}{2p}\right)$

- vi) adaptive SLOPE - as in point iv) but the in final stage use weighted SLOPE with BH sequence at FDR level 0.2.
  - vii) extra 5 points - adaptive Bayesian SLOPE at FDR=0.2.
- a) For methods iii)-vii) estimate FDR and power.
  - b) For all methods apart from iii) estimate the mean square errors of the estimators of  $\beta$  and  $\mu = X\beta$ .
2. Repeat Problem 1 when  $X_{i.} \sim N(0, \frac{1}{n}\Sigma)$ , where  $\Sigma_{ii} = 1$  and  $\Sigma_{ij} = 0.5$ .

Malgorzata Bogdan