

# Sparse Modelling: Best Subset and the Lasso

## Exercises

### 1 CASI

#### Chapter 16 Exercises

1. In forward-stepwise regression, we include the variable at each step that improves the residual-sum-of squares the most. You notice that in a software package you were using, the variable is chosen that has the maximum absolute correlation with the current residual. Are these two approaches equivalent? Explain.
2. Describe in some detail an efficient approach for computing the forward-stepwise regression model path.
4. Run a simulation to compare the df of best-subset regression and lasso. Use  $p = 30$  variables and  $n = 200$  observations to build an  $\mathbf{X}$  matrix, generated from a multivariate Gaussian distribution with non-trivial covariance (of your choice). Now pose a response model  $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$  and specify  $\beta$  in advance. In your simulations hold  $\mathbf{X}$  and  $\beta$  fixed, and generate new  $\boldsymbol{\varepsilon}$  at each run. Make a plot similar to the right plot in Figure 16.8