

Model selection

Best subsets regression

Statement of problem

- A common problem is that there is a large set of candidate predictor variables.
- Goal is to choose a small subset from the larger set so that the resulting regression model is **simple**, yet have **good predictive ability**.

Example: Cement data

- Response y : heat evolved in calories during hardening of cement on a per gram basis
- Predictor x_1 : % of tricalcium aluminate
- Predictor x_2 : % of tricalcium silicate
- Predictor x_3 : % of tetracalcium aluminato ferrite
- Predictor x_4 : % of dicalcium silicate