Regularized regression I

Motivation

Selecting features for prediction

- Prediction problems might involve data with many potential features (and less domain knowledge)
  - Contrasts theory-based hypothesis testing perspective
  - Relates to building parsimonious models (e.g. with respect to AIC, BIC)
- Interest in developing sparse models to improve:
  - Prediction accuracy
  - Model interpretability

$\rightarrow$ Remove non-informative variables to improve effectiveness

Motivation

Is feature selection necessary?

- Regression models
  - Run into problems in high dimensions (large p, small n)
    - $p \approx n$: Overfitting and poor prediction performance
    - $p > n$: OLS can not be estimated
- Tree-based models
  - Involve built-in feature selection
  - Less affected by irrelevant features
- Support vector machines, neural networks
  - Negatively affected by irrelevant features

Selecting Predictors

Figure: Consequences of non-informative predictors
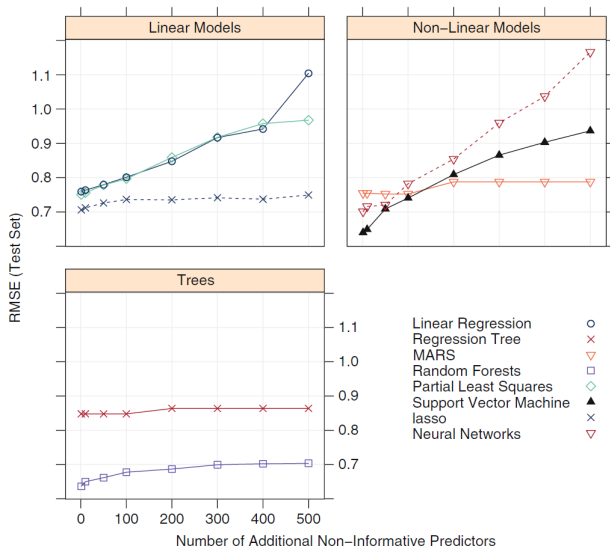
Feature Selection Methods

Feature selection methods

- Wrapper
    - Search algorithms that add and/or remove predictors to optimize performance
    - e.g. forward, backward, and best subset selection
- Filter
    - Test individual predictors outside of the predictive model
    - e.g. t-tests, r, χ^2
- ℓ_1 **regularization**