Concepts in Machine Learning
Winter Institute in Data Science

Ryan T. Moore

8 January 2020

How do we build models?

How do we build models?

What are our goals?

How do we build models?

What are our goals?

1. Generative modeling
2. Predictive modeling

How do we build models?

How do we build models?

- Theory
  (novel theory, prior theory, prior findings)

How do we build models?

- Theory
  (novel theory, prior theory, prior findings)
- Raw data
  ("data look nonlinear, so $\ldots + x^2 + \ldots$")

How do we build models?

- Theory
  (novel theory, prior theory, prior findings)
- Raw data
  ("data look nonlinear, so $\ldots + x^2 + \ldots$")
- Specification searching
  (repeat modeling with same data)

How do we build models?

- Theory
  (novel theory, prior theory, prior findings)
- Raw data
  ("data look nonlinear, so $\ldots + x^2 + \ldots$")
- Specification searching
  (repeat modeling with same data)
- Testing and training
  (repeat modeling, different data)

What to include, when thousands of predictors?

- “machine learning”
- “data mining”
- “. . . ”

Feature Selection

- Wrappers: pick subset of covars, train on data (estimate model), test on hold-out, score predictions. Keep best-scoring subset.

Feature Selection

- Wrappers: pick subset of covars, train on data (estimate model), test on hold-out, score predictions. Keep best-scoring subset.
- Filters: correlate covars with outcome. Keep strongest.

Feature Selection

- Wrappers: pick subset of covars, train on data (estimate model), test on hold-out, score predictions. Keep best-scoring subset.
- Filters: correlate covars with outcome. Keep strongest.
- Embeds: select features and estimate model at same time. Penalize using more predictors.

Embedded Regularization Methods

OLS reminder

Minimize SSR:

$$\text{argmin}(\beta) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\text{argmin}(\beta) \sum_{i=1}^{n} \left(y_i - \mathbf{X}\hat{\beta}\right)^2$$

Embedded Regularization Methods

L1 regularization: the LASSO (Least Absolute Shrinkage and Selection Operator)

$$\text{argmin}(\beta)\left[\sum_{i=1}^{n}\left(y_i - \mathbf{X}\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{k}|\hat{\beta}_j|\right]$$

Embedded Regularization Methods

L1 regularization: the LASSO (Least Absolute Shrinkage and Selection Operator)

$$\text{argmin}(\beta)\left[\sum_{i=1}^{n}\left(y_i - \mathbf{X}\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{k}|\hat{\beta}_j|\right]$$

L2 regularization: Ridge regression

$$\text{argmin}(\beta)\left[\sum_{i=1}^{n}\left(y_i - \mathbf{X}\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{k}\hat{\beta}_j^2\right]$$

Embedded Regularization Methods

Mix L1 and L2: Elastic net

$$\operatorname{argmin}(\beta)\left(\frac{\sum_{i=1}^{n}\left(y_i-\mathbf{X}\hat{\beta}\right)^2}{2n}+\lambda\left[\alpha\sum_{j=1}^{k}|\hat{\beta}_j|+\frac{1-\alpha}{2}\sum_{j=1}^{k}\hat{\beta}_j^2\right]\right)$$

Embedded Regularization Methods

Mix L1 and L2: Elastic net

$$\text{argmin}(\beta)\left(\frac{\sum_{i=1}^{n}\left(y_i - \mathbf{X}\hat{\beta}\right)^2}{2n} + \lambda\left[\alpha\sum_{j=1}^{k}|\hat{\beta}_j| + \frac{1-\alpha}{2}\sum_{j=1}^{k}\hat{\beta}_j^2\right]\right)$$

Regularized trees, ...

R packages for Regularization

- `glmnet`
- `caret`

See also `tidymodels`, `parsnip`, ...