Model selection: model order and regularization

Fraida Fund

Contents

In this lecture

- Bias-variance tradeoff
- Model selection
- Cross validation
- Regularization

Model selection

Occam's razor

Choosing model complexity

- Model order selection
- Feature selection
- Model class selection

Model selection problem

TODO - change this to feature selction? Use Andrew Ng notes http://cs229.stanford.edu/notes/cs229-notes5.pdf

- Linear model: $\hat{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d$
- Model target y as a function of features $\mathbf{x} = (x_1, \cdots, x_d)$
- Many features, only some are relevant
- Model selection problem: fit a model with a small number of features

Model selection problem - formal

Problem: determine a subset of features $I \subseteq 1, \cdots, d$ with $|I|$ small.

Fit model

$$\hat{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d$$

where $\beta_j = 0$ for all $j \notin I$

Motivation for model selection problem

- Limited data
- Very large number of features
- Examples: spam detection using "bag of words", EEG, DNA MicroArray data

Cross validation

Avoiding data leakage in CV

Regularization

Penalty for model complexity

With no bounds on complexity of model, we can always get a model with zero training error on finite training set - overfitting.

Regularization vs. standard LS

Least squares estimation:

$$\hat{\beta} = \operatorname*{arg\,min}_{\beta} RSS(\beta), \quad RSS(\beta) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Regularized estimation with a **regularizing function** $\phi(\beta)$:

$$\hat{\beta} = \operatorname*{arg\,min}_{\beta} J(\beta), \quad J(\beta) = RSS(\beta) + \phi(\beta)$$

Common regularizers: Ridge and LASSO

Ridge regression (L2):

$$\phi(\beta) = \alpha \sum_{j=1}^{d} |\beta_j|^2$$

LASSO regression (L1):

$$\phi(\beta) = \alpha \sum_{j=1}^{d} |\beta_j|$$