Applied Machine Learning

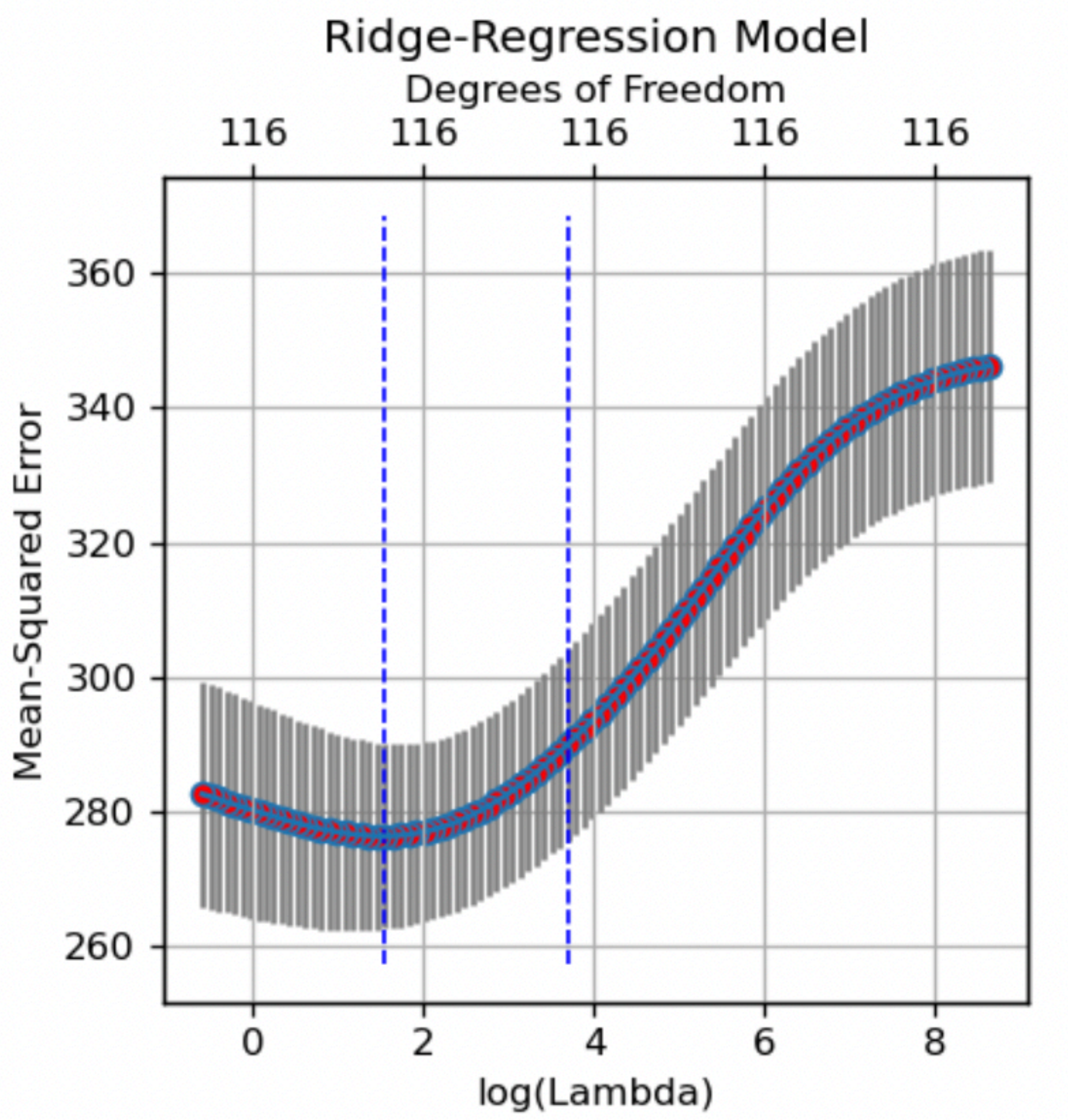
Linear Regression Models

Linear Regression Models

- Explanatory Variables
- Bias and Variance
- Strategies to choose models

Explanatory variables

- What explanatory variables to choose?
  - Some may not be as important
  - Some may be missing
- Predictions may be affected by
  - Irreducible error
  - Error due to bias
  - Error due to variance



Bias and Variance in Regularization

- $Y = f(\mathbf{x}) + \xi$
  - $\mathbf{x}$: vector of explanatory variables
  - Y: random dependent variable
  - $f(\cdot)$: estimation of dataset model
  - ξ: noise
- expected value of error in prediction: $\mathbb{E}[(Y - \hat{f}(x))^2]$
  - $\hat{f}(\cdot)$: estimation of $f(\cdot)$

- best possible $\hat{f}(\cdot)$
  - perform well on all the possible data (not only on a chosen subset)
  - achieves same predictions as $\mathbb{E}[\hat{f}]$

Bias and Variance in Regularization

- $Y = f(\mathbf{x}) + \xi$
  - $\mathbf{x}$: vector of explanatory variables
  - Y: random dependent variable
  - $f(\cdot)$: estimation of dataset model
  - ξ: noise
- expected value of error in prediction: $\mathbb{E}[(Y - \hat{f}(x))^2]$
  - $\hat{f}(\cdot)$: estimation of $f(\cdot)$

- $\mathbb{E}[\xi] = 0$
- $\text{var}(Y) = \text{var}(\xi) = \sigma_\xi^2$
- $\mathbb{E}[(Y - \hat{f}(x))^2] = \sigma_\xi^2 + (f - \mathbb{E}[\hat{f}])^2 + \text{var}(\hat{f})$
  - best model: $\mathbb{E}[\hat{f}]$
  - prediction: $\hat{f}$

Bias and Variance in Regularization

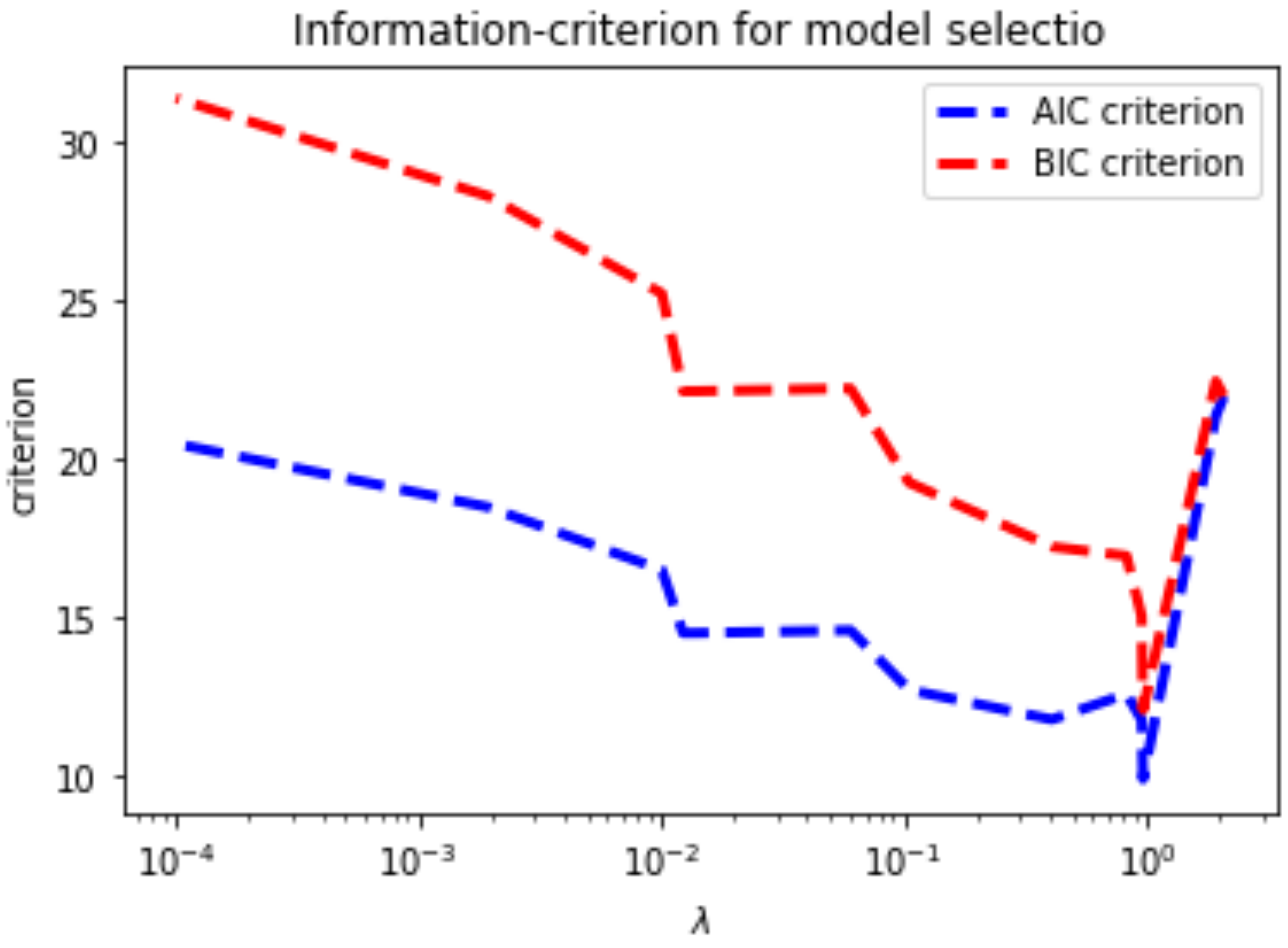
- $\mathbb{E}[(Y - \hat{f}(x))^2] = \sigma_\xi^2 + (f - \mathbb{E}[\hat{f}])^2 + \text{var}(\hat{f})$
  - σ_ξ^2: irreducible error
  - $(f - \mathbb{E}[\hat{f}])^2$: bias
  - $\text{var}(\hat{f})$: variance
- Choose model parameters
  - small enough to control variance
  - large enough to prevent bias
- Simple models
  - Easy to estimate best parameters: low variance
  - More specific: high bias
- More complex models
  - Harder to estimate best parameters: high variance
  - More general model: low bias

Choosing Models

- Measure Information criteria to find good model sizes
  - Akaike Information Criterion (AIC)
  - Bayes Information Criterion (BIC)
- Penalty that increases with the number of parameters

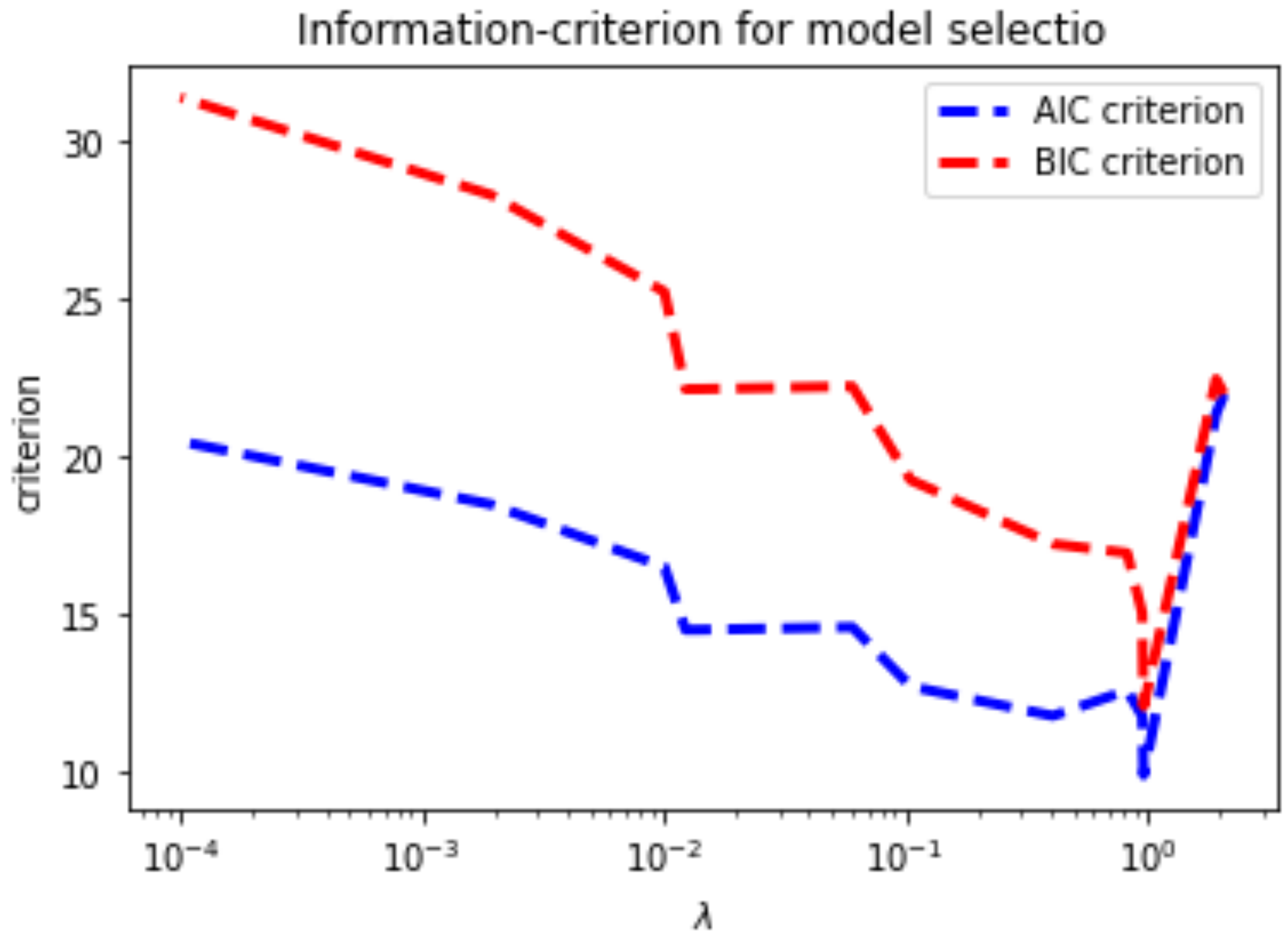
Information Criteria

- Model
    - $k = d + 1,\ d$ explanatory variables
    - $\mathscr{L}$: log-likelihood
- Akaike Information Criterion (AIC)
    - $2k - 2\mathscr{L}$
- Bayes Information Criterion (BIC)
    - $2k \log N - 2\mathscr{L}$



Cross-Validation to Choose Models

- F folds:
  - "leave-one-out": F=N
- for each combination of parameters
  - iteratively with F-1 folds as cross-validation training set
    - find β on Cross-Validation Train Set
    - MSE on held-out Validation set
  - error on held-out data is average on all iterations



Exploration of Parameters

- Forward stagewise regression
  - working set = $\{\phi\}$
  - iterate until no improvement
    - for each explanatory variable not in working set
      - add variable to model, evaluate, keep best score
    - if best score better than score for working set, add variable to working set

Exploration of Parameters

- Backward stagewise regression
    - working set = all the explanatory variables
    - iterate until no improvement
        - for each explanatory variable in working set
            - evaluate without variable, evaluate, keep best score
        - if best score better than score for working set, remove variable from working set

Variable Importance

- Regression coefficients value does not reflect importance
  - Have same units of corresponding explanatory variable
  - Statistical significance
  - Even low coefficients have impact on regression

Linear Regression Models

- Explanatory Variables
- Bias and Variance
- Strategies to choose models

Applied Machine Learning

Linear Regression Models