

Nonparametric Regression for House Price Prediction

MMATH STAT 844 Presentation

Yuqi Jing

Background & Methods

- Why Nonparametric Regression?

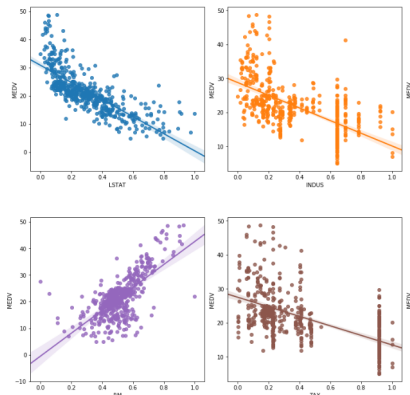


Figure: The response **medv** plotted against 4 predictors including lstat, indus, rm and tax. The Boston dataset contains the median value (medv) and 12 sociodemographic statistics values of the Boston houses.

Background & Methods

- Why Nonparametric Regression?
 - Housing Price Prediction is Hard!
 - **Boston Housing Dataset:** the house price is nonlinear in several predictors.

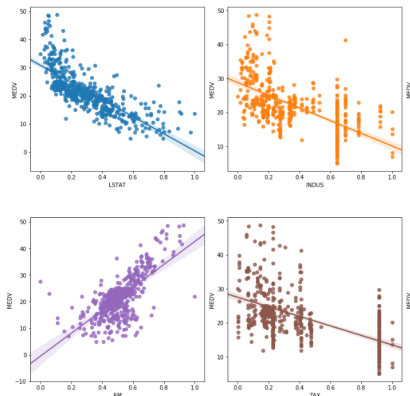


Figure: The response **medv** plotted against 4 predictors including lstat, indus, rm and tax. The Boston dataset contains the median value (medv) and 12 sociodemographic statistics values of the Boston houses.

Background & Methods

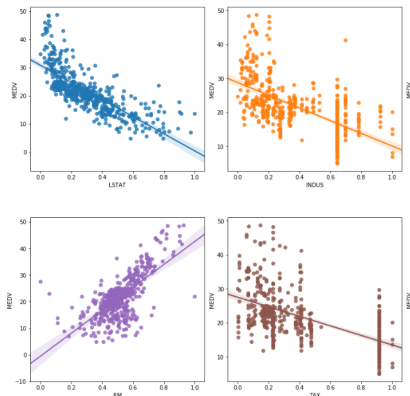


Figure: The response **medv** plotted against 4 predictors including lstat, indus, rm and tax. The Boston dataset contains the median value (medv) and 12 sociodemographic statistics values of the Boston houses.

- Why Nonparametric Regression?
 - Housing Price Prediction is Hard!
 - **Boston Housing Dataset:** the house price is nonlinear in several predictors.
- Smoothing Regression
 - $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$, where \mathbf{S}_λ is *smoother matrix*.

Background & Methods

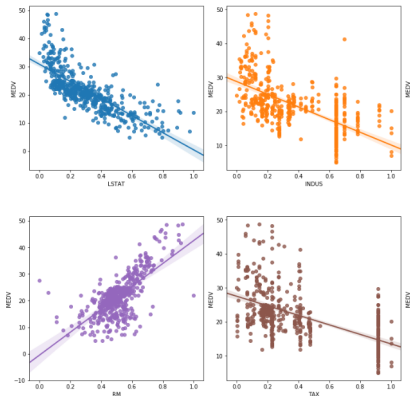


Figure: The response **medv** plotted against 4 predictors including **lstat**, **indus**, **rm** and **tax**. The Boston dataset contains the median value (**medv**) and 12 sociodemographic statistics values of the Boston houses.

- Why Nonparametric Regression?
 - Housing Price Prediction is Hard!
 - **Boston Housing Dataset:** the house price is nonlinear in several predictors.
- Smoothing Regression
 - $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$, where \mathbf{S}_λ is *smoother matrix*.
 - Smoothing Spline
 - Smoothing Kernel

Background & Methods

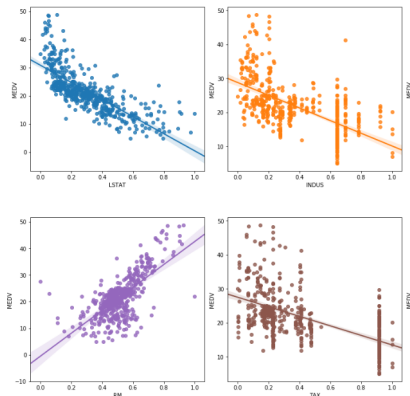


Figure: The response **medv** plotted against 4 predictors including lstat, indus, rm and tax. The Boston dataset contains the median value (medv) and 12 sociodemographic statistics values of the Boston houses.

- Why Nonparametric Regression?
 - Housing Price Prediction is Hard!
 - **Boston Housing Dataset:** the house price is nonlinear in several predictors.
- Smoothing Regression
 - $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$, where \mathbf{S}_λ is *smoother matrix*.
 - Smoothing Spline
 - Smoothing Kernel
- Reproducing Kernel Method
 - $\hat{\mathbf{f}} = \mathbf{K} \hat{\boldsymbol{\alpha}}$, where \mathbf{K} is *kernel matrix*.
 - SVM for regression

Background & Methods

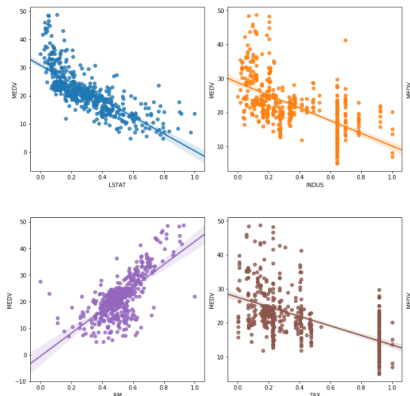
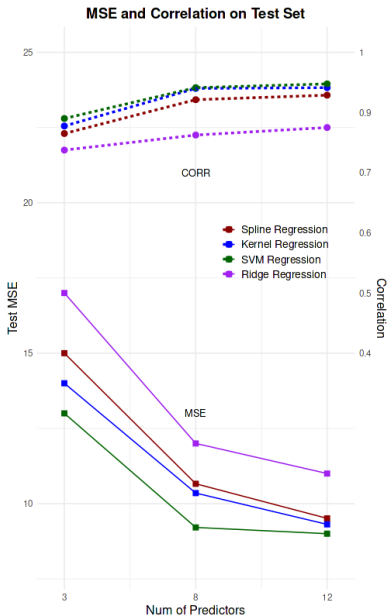


Figure: The response **medv** plotted against 4 predictors including lstat, indus, rm and tax. The Boston dataset contains the median value (medv) and 12 sociodemographic statistics values of the Boston houses.

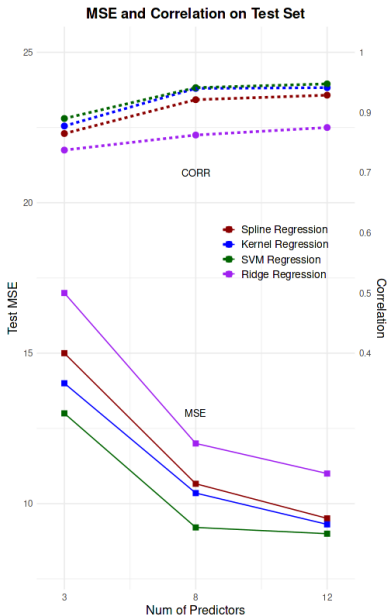
- Why Nonparametric Regression?
 - Housing Price Prediction is Hard!
 - **Boston Housing Dataset:** the house price is nonlinear in several predictors.
- Smoothing Regression
 - $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$, where \mathbf{S}_λ is *smoother matrix*.
 - Smoothing Spline
 - Smoothing Kernel
- Reproducing Kernel Method
 - $\hat{\mathbf{f}} = \mathbf{K} \hat{\boldsymbol{\alpha}}$, where \mathbf{K} is *kernel matrix*.
 - SVM for regression

Experiments & Results



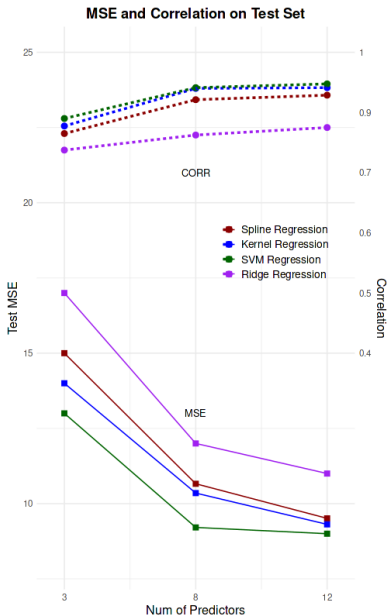
- Randomly split the dataset into training set (80%) and test set (20%).
 - Fit using num of predictors as 3, 8 and 12 respectively selected by LASSO.

Experiments & Results



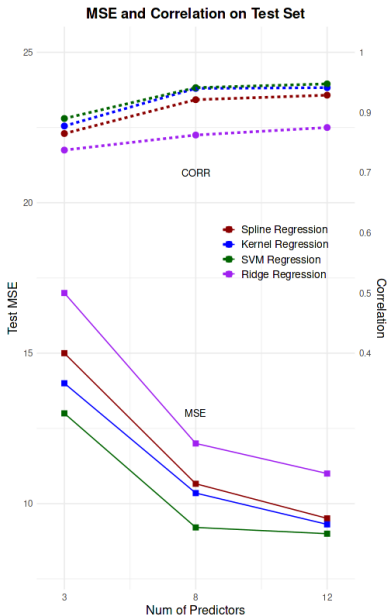
- Randomly split the dataset into training set (80%) and test set (20%).
 - Fit using num of predictors as 3, 8 and 12 respectively selected by LASSO.
 - *lstat*, *rm*, *ptratio* are three most important predictors.

Experiments & Results



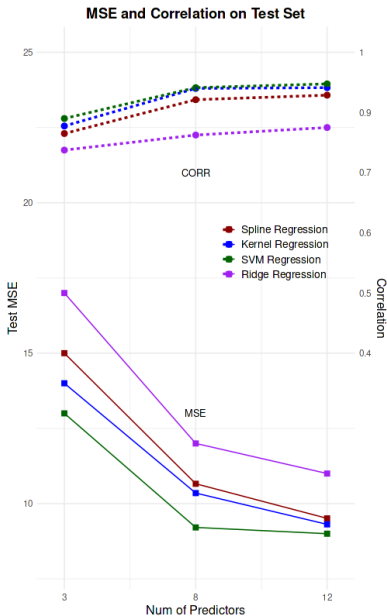
- Randomly split the dataset into training set (80%) and test set (20%).
 - Fit using num of predictors as 3, 8 and 12 respectively selected by LASSO.
 - *lstat*, *rm*, *ptratio* are three most important predictors.
 - Repeat the experiments 100 times.
- SVM Regression provides lowest testing error and highest correlation among three methods.

Experiments & Results



- Randomly split the dataset into training set (80%) and test set (20%).
 - Fit using num of predictors as 3, 8 and 12 respectively selected by LASSO.
 - *lstat*, *rm*, *ptratio* are three most important predictors.
 - Repeat the experiments 100 times.
- SVM Regression provides lowest testing error and highest correlation among three methods.
 - Ridge provided as the baseline.

Experiments & Results



- Randomly split the dataset into training set (80%) and test set (20%).
 - Fit using num of predictors as 3, 8 and 12 respectively selected by LASSO.
 - **lstat**, **rm**, **ptratio** are three most important predictors.
 - Repeat the experiments 100 times.
- SVM Regression provides lowest testing error and highest correlation among three methods.
 - Ridge provided as the baseline.
 - **Conclusion**: SVM Regression will be selected to predict the house price.