Revision

April 7, 2007

1 Basic nonparametric smoothing methods

To explore the (complicated) conditional expectation relationship between two variables, say $\mathbf{x}$ and Y, in model

$$Y = g(\mathbf{x}) + \varepsilon$$

where $E(\varepsilon|\mathbf{x}) = 0$, we have introduced two methods: polynomial spline and kernel smoothing (Nadaraya-Watson kernel method and local linear smoothing method).

Suppose we have observations $(\mathbf{x}_1, Y_1), ..., (\mathbf{x}_n, Y_n)$. The basics we need to know

1. the basic idea of the methods

2. The estimation of $g(x)$ using the two methods and their expressions.

3. The confidence interval for $g(x)$

4. How to apply the method to the estimation of more complicated models.

    For example, how to estimate the varying coefficient model using NW kernel estimation or polynomial splines.

5. How to calculate the CV values of the complicated models.

    For example, how to calculate the CV values of the partially linear regression model, varying coefficient models and other models.

6. How to calculate the confidence band for functions in the complicated models.

2 CV value criterion

How to calculate CV and GCV values in all of the models? The basic idea of CV.

3   Models and methods

The models and methods we discussed are

1. The partially linear regression model

2. The single-index model

3. The Projection pursuit regression (model)

4. The varying coefficient model

5. The generalize additive model

6. Multivariate Adaptive regression splines (MARS)

7. Classification and regression tree (CART)

8. Support vector machine learning

9. Flexible Discriminant analysis

10. k-Nearest-Neighbor classification

11. Boosting method

You need to know

1. what kind of problems the above models or methods can be applied

2. the basic idea of estimating the models and implementation of the methods

3. For some models, how to interpret the calculation results

4   Codes and libraries

1. For basic smoothing methods:

   (a) kernel methods: ks.R, cvh.R, ksm.R

   (b) spline methods: pspline.R; gam

   (c) ridge regression.

2. For statistical models

    (a) linear regression model: lm;

    (b) partially linear regression model: plr.R, gam

    (c) varying coefficient model: vcm.R

    (d) Single-index model and projection pursuit regression: sim.R; ppr.

    (e) generalized additive model: gam

    (f) multi-dimensional model: tree; mars [in mda]; svm [in e1071];

3. For classification

    tree; svm; fda [in mda]; kknn; adaboost [in boost]