

Lecture 11

Jussi Klemelä

April 6, 2009

1 Choice of the regression method

Nonparametric methods are flexible and able to adapt to various kinds of data but they suffer from the curse of dimensionality and the lack of interpretability. Semiparametric methods are able to cope with quite high dimensional data and they are easier to interpret, but they are less flexible and their use may result to large modeling errors.

1.1 Interpretation of regression models

We recall regressograms. A regression tree is a regressogram where the partition is obtained by a recursive partition. This makes it possible to visualize the regressogram as a tree. Much of the popularity of regression trees comes from this effective visualization of regression functions as binary trees. Figure 1 shows a regression tree based on the Boston housing data. The data was gathered for Harrison and Rubinfeld (1978) to study the effect of pollution to housing values. The regression tree of the figure was grown in Belsley, Kuh and Welsch (1980), but we have printed Figure 8.1 of Section 8.2 of Breiman, Friedman, Olshen and Stone (1984). There are 14 variables and 506 census tracts. The median value of homes is the response variable and the other 13 variables are explanatory variables. Only four of the variables appear in the regression tree: average number of rooms (RM), percent lower-status population (LSTAT), weighted distance to employment centers (DIS), and crime rate (CRIM). In particular, the air pollution variable does not appear.

1.2 The quality of fit

We divide the sample $(X_1, Y_1), \dots, (X_n, Y_n)$ to the estimation part and to the test part: $(X_1, Y_1), \dots, (X_{n^*}, Y_{n^*})$ and $(X_{n^*+1}, Y_{n^*+1}), \dots, (X_n, Y_n)$, where

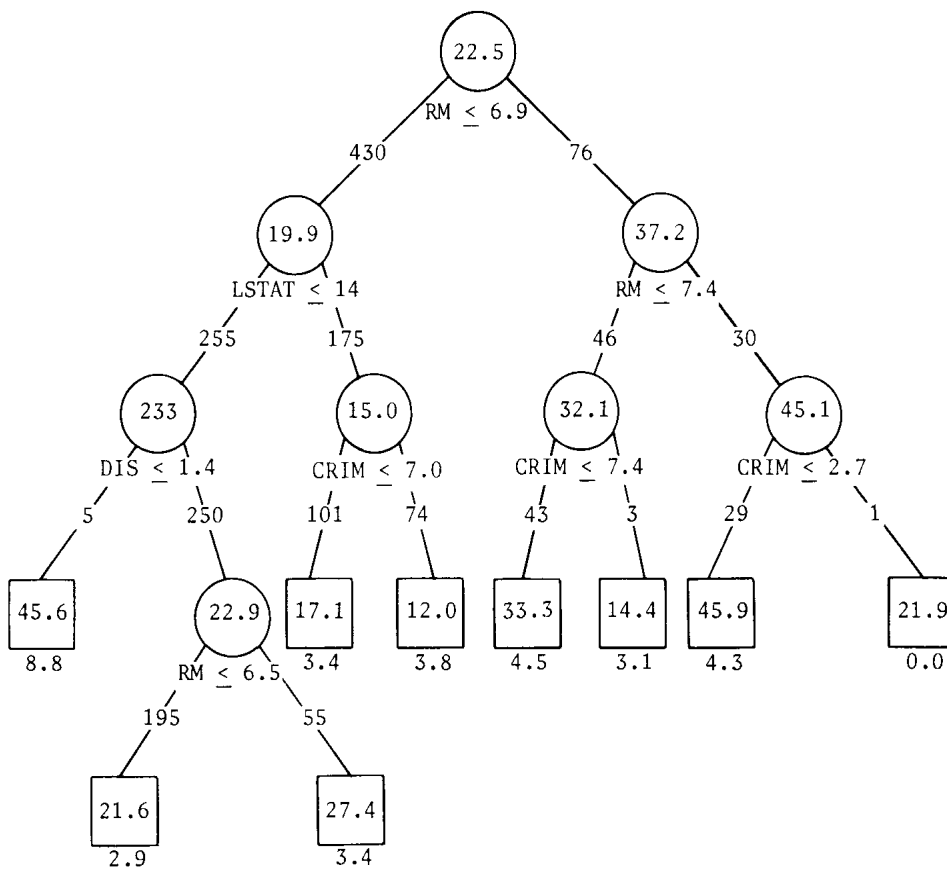


FIGURE 8.1

Figure 1: A regression tree from the Boston housing data.

$n^* = \lfloor n/2 \rfloor$. The estimation part is used to construct regression function estimators $\hat{f}_1, \dots, \hat{f}_L$. The testing part is used to calculate the indicators for the quality of fit:

$$SSE_l = \sum_{i=n^*+1}^n \left(Y_i - \hat{f}_l(X_i) \right)^2, \quad l = 1, \dots, L.$$

Finally we choose the method \hat{f}_l minimizing SSE_l .

1.3 Smoothing parameter selection in kernel estimation

The kernel estimator $\hat{f} = \hat{f}_h$ depends on the smoothing parameter $h > 0$. We can use data based methods for choosing the smoothing parameter. The smoothing parameter of kernel estimator can be chosen by minimizing an estimator of the mean integrated squared error as a function of h . The mean integrated squared error is defined as

$$\text{MISE}(h) = E \int_{\mathbf{R}^d} \left(\hat{f}_h(x) - f(x) \right)^2 w(x) f_X(x) dx.$$

An estimator of $\text{MISE}(h)$ could be chosen as

$$\frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{f}_h(X_i) \right)^2 w(X_i).$$

However, this estimator can be made arbitrarily small by letting $h \downarrow 0$. Let us define the leave-one-out kernel estimator

$$\hat{f}_{h,-i}(x) = \sum \{ p_{j,-i}(x) Y_j : j = 1, \dots, n, j \neq i \},$$

where

$$p_{j,-i}(x) = \frac{p_j(x)}{\sum \{ p_j(x) : j = 1, \dots, n, j \neq i \}}, \quad j = 1, \dots, n, j \neq i.$$

Then the estimator of $\text{MISE}(h)$ based on cross-validation is defined as

$$\text{MISE}_n(h) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{f}_{h,-i}(X_i) \right)^2 w(X_i).$$

It can be proved that this is an unbiased estimator of $\text{MISE}(h)$. Note that

$$\text{MISE}_n(h) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{f}_h(X_i) \right)^2 (1 - p_i(X_i))^{-2} w(X_i).$$

Indeed,

$$\text{MISE}_n(h) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{f}_h(X_i) \right)^2 \left(\frac{Y_i - \hat{f}_{h,-1}(X_i)}{Y_i - \hat{f}_h(X_i)} \right)^2 w(X_i)$$

and

$$\begin{aligned} \frac{Y_i - \hat{f}_h(X_i)}{Y_i - \hat{f}_{h,-1}(X_i)} &= \frac{Y_i \sum_j p_j(X_i) - \sum_j p_j(X_i) Y_j}{Y_i \sum_{j \neq i} p_{j,-i}(X_i) - \sum_{j \neq i} p_{j,-i}(X_i) Y_j} \\ &= \frac{Y_i \sum_j p_j(X_i) - \sum_j p_j(X_i) Y_j}{Y_i \sum_{j \neq i} p_j(X_i) - \sum_{j \neq i} p_j(X_i) Y_j} \times \sum_{j \neq i} p_j(X_i) \\ &= 1 - p_i(X_i), \end{aligned}$$

because $\sum_{j \neq i} p_j(X_i) = 1 - p_i(X_i)$. A penalizing function approach minimizes

$$G(h) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{f}_h(X_i) \right)^2 P(p_i(X_i)) w(X_i),$$

where $P : (0, \infty) \rightarrow \mathbf{R}$ is a penalizing function. We have the following choices:

1. $P(u) = (1 - u)^{-2}$,
2. $P(u) = 1 + 2u$,
3. $P(u) = \exp(2u)$,
4. $P(u) = (1 + u)/(1 - u)$,
5. $P(u) = (1 - 2u)^{-1}$.

2 Portfolio selection

2.1 The selection of the method

To illustrate the process of regression analysis we use the example of portfolio selection, where the following choices have to be made.

- Choice of the explanatory variables
 - We can use the previous price relatives. How many previous price relatives we should choose?

- We can add price relatives of other assets.
- We can use technical indicators instead of plain price relatives
- Choice of the transformation of the explanatory variables.
 - We can use copula transform.
 - We can use data sphering.
 - We can use a combination of these
- Choice of the response variable.
 - We can use CRRA utility functions
 - We can use a return-risk optimization. (This takes us to the estimation of conditional variance or conditional quantiles and conditional shortfall.
- Choice of the regression method: kernel, nearest neighborhood, regressogram, . . .
 - Choice of the smoothing parameter.
 - Choice of the other parameters of the estimate (kernel, the minimal number of observations in a regressogram neighborhood).
- Choice of the assets.
 - We can include bank account.
 - Stock indices, stocks, currencies, bonds, commodities.
- The range of the portfolio parameters.
 - Long only portfolios.
 - “All-in-one-asset-at-a-time” allocation.
 - Allowing shorting.

2.2 Combination of strategies

Typically we have to choose the method from several alternatives. For example, when we use the kernel method we have to choose the smoothing parameter. Also, the explanatory variables can be chosen many ways and we can consider each way of choosing the explanatory variables as its own method. Suppose we have L methods which have produced portfolio vectors

$$b_0^{(l)}, \dots, b_{T-1}^{(l)}, \quad l = 1, \dots, L.$$

We have the corresponding L wealth processes

$$W_T^{(l)} = W_0 \prod_{t=0}^{T-1} b_t^{(l)} \cdot U_{t+1}.$$

We can choose weights $q_t^{(l)}$, $l = 1, \dots, L$, for the L strategies, where the weights satisfy $q_t^{(l)} > 0$, $\sum_{l=1}^L q_t^{(l)} = 1$. The combined portfolio vectors are

$$B_t = \sum_{l=1}^L q_t^{(l)} b_t^{(l)}, \quad t = 0, \dots, T-1.$$

Clearly $B_t^i > 0$, $\sum_{i=1}^d B_t^i = 1$. A useful way to choose the weights is to take

$$q_t^{(l)} = \frac{W_{t-1}^{(l)}}{\sum_{l=1}^L W_{t-1}^{(l)}}, \quad t = 1, 2, \dots \quad (1)$$

and the initial weights $q_0^{(l)}$, $l = 1, \dots, L$, by some rule, for example

$$q_0^{(l)} = \frac{1}{L}, \quad l = 1, \dots, L.$$

This strategy weights more the successful strategies. Note however that this strategy is equivalent to distributing the wealth according to the initial weights $q_0^{(l)}$ and running the strategies independently. The wealth obtained by using portfolio weights B_0, \dots, B_{T-1} with the weighting rule (1) is the combination of the wealths $W_T^{(l)}$ according to the initial weights:

$$W_T(B) = \sum_{l=1}^L q_0^{(l)} W_T^{(l)}.$$

(We are not constantly rebalancing among the L strategies.)

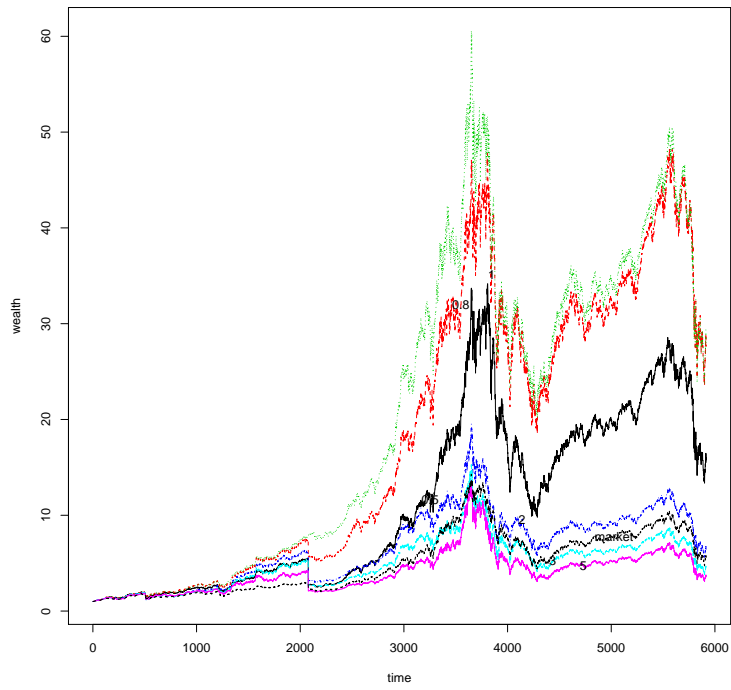


Figure 2: SP500-Nasdaq, 20090326, $k=5$, copula transform with Gaussian marginals, kernel estimator with $h = 0.5, 0.8, 1, 2, 3, 5$, Gaussian kernel.

2.3 Example

Figure 2 shows the case with $k = 5$, $h = 0.5, 0.8, 1, 2, 3, 5$, Gaussian kernel. We have used the standardization of the marginals to be Gaussian.

3 Examination

Possible questions in the examination:

- 12) Define the method of cross-validation to choose the smoothing parameter in kernel regression.

References

- Belsley, D. A., Kuh, A. and Welsch, R. E. (1980), *Regression diagnostics*, Wiley, New York.
- Breiman, L., Friedman, J., Olshen, R. and Stone, C. J. (1984), *Classification and Regression Trees*, Chapman & Hall, New York.
- Harrison, D. and Rubinfeld, D. L. (1978), 'Hedonic prices and the demand for clean air', *J. Envir. Econ. and Management* **5**, 81–102.