Given TS $(x_1, y_1) \dots (x_n\ y_n)$ iid obs of (X, Y)

Goal Make prediction rule $f(x)$

Optimal predictor $f(x) = E(Y \mid x)$

Turbo

Approximate $E(Y|x)$ by piecewise linear spline

Dictionary

$B_j(x) = (x - x_j)_+ \quad j = 1 \dots n-1$

$B_n(x) = 1$

Fit sequence of models using stepwise forward selection

$$j_1 = \underset{j}{\mathrm{argmin}}\ \underset{a}{\mathrm{argmin}} \sum_{i=1}^{n} (y_i - a B_j(x_i))^2 \qquad (*)$$

$$j_2 = \underset{j}{\mathrm{argmin}}\ \underset{\underline{a}}{\mathrm{argmin}} \sum_{i=1}^{n} (y_i - a_1 B_{j_1} - a_2 B_j)^2$$

$$\vdots$$

etc

Question: How to pick the best # of basis functions?

Several options

m-fold cross-validation

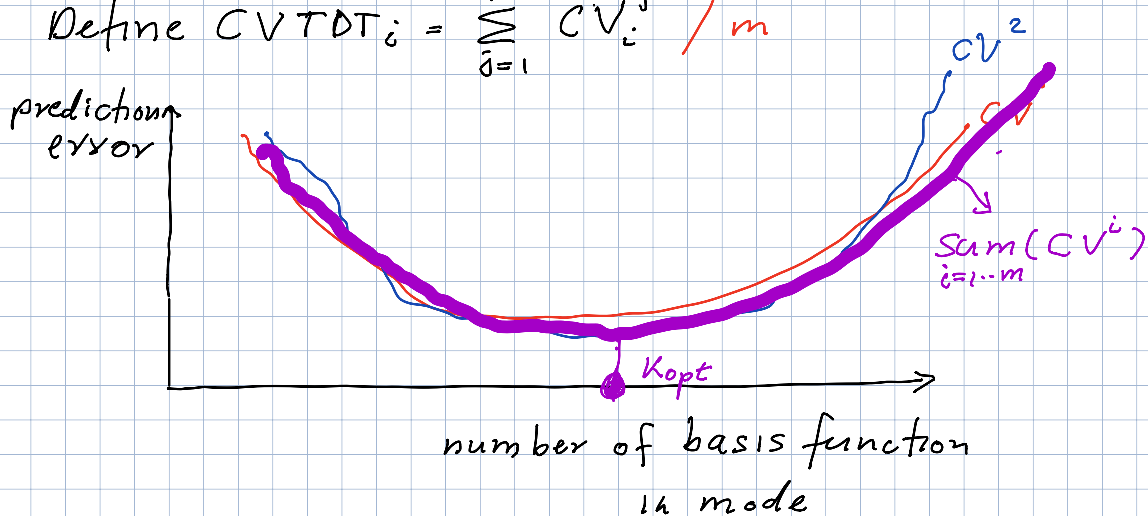
(1) Randomly divide data set S into m equally sized chunks $S_1 \dots S_m$ $\quad n - \frac{n}{m}$ obs

(2) Apply stepwise selection to S/S_1, obtaining model $g_1^1(x) \dots \dots g_n^1(x)$

(3) Use each of these models to predict obs in S_1 obtaining $CV_1^1 \dots CV_n^1$

(4) Repeat (2) (3) for $S_2 \dots S_m$ obtaining $g_1^1 \dots g_n^1 \quad g_1^2 \dots g_n^2 \quad CV_1^m \dots CV_n^m$

(5) Define $CVTOT_i = \sum_{j=1}^{m} CV_i^j \,/\, m$



(6) Choose $K = \underset{i}{\operatorname{argmin}}\, (CVTOT_i)$

Leave-one-out cross-validation LOOCV

Recall Kernel smoothing

$$\hat{y} = W y \qquad \hat{y}_i = \sum_j w_{ij} y_j \Big/ \sum_j w_{ij}$$

Define $w_{ij}^* = w_{ij} \Big/ \sum_j w_{ij} \qquad \sum_j w_{ij}^* = 1$

$$\hat{y}_i = \sum_j w_{ij}^* y_j$$

$\hat{y}_i^{-i}$ predicted response for training obs i computed without using obs i

$$\hat{y}_i^{-i} = \sum_{j \neq i} w_{ij}^* y_j \Big/ (1 - w_{ii}^*)$$

$$r_i^{-i} = (y_i - \hat{y}_i^{-i}) = \frac{y_i - \sum_{j \neq i} w_{ij}^* y_i}{(1 - w_{ii}^*)} = \frac{y_i - \hat{y}_i}{(1 - w_{ii}^*)}$$

CV residual

$$CV = \frac{1}{n} \sum (r_i^{-i})^2 = \frac{1}{n} \sum \frac{r_i^2}{(1 - w_{ii}^*)^2}$$

For least squares

$$\hat{y} = H y \qquad H \text{ hat matrix}$$

One can show that $r_i^{-i} = r_i / (1 - h_{ii})$

$$CV = \frac{1}{n} \sum \frac{r_i^2}{(1 - h_{ii})^2}$$

"Generalized" Cross-validation for least squares

Replace the h_{ii} by their average
trace $(H)/n$

For a least squares model with K predictors
trace $(H) = K$

$$GCV = \frac{1}{n} \sum r_i^2 \Big/ \left(1 - \frac{K}{n}\right)^2$$

linear smoother $\hat{y} = \underset{\sim}{W} \underset{\sim}{y}$

elements of W can depend on $x_1, \ldots x_n$, but not on y's

$f(x) = \langle \underline{w}, \underset{\sim}{y} \rangle$

$w(x, x_1 \ldots x_n)$