Expansion based Part 5

Given TS $(x_1, y_1) \dots (x_n, y_n)$ $x_1 < x_2 \quad < x_n$

Goal Estimate $f(x) = E(Y|x)$

Turbo Approximate $f(x)$ by order 2 spline (continuous, piece wise linear)

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

$$B_n(x) = 1$$

Options:

- Fewer basis function.
- Higher order splines (cubic spline)

To construct approximation, we fit a sequence of models using stepwise forward selection

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

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

$$\vdots \text{ etc}$$

Options:

- Backward elimination
- Forward / backward mixture
- Best subsets only feasible for small dictionaries

Key question: How to pick # K of basis functions

Authors use generalized cross-validation

RSS_K residual sum of squares for model with K basis functions.

Choose $\hat{K} = \underset{K}{\operatorname{argmin}} \; RSS_K \,/(1-3\frac{K}{n})^2$ *

$K \nearrow \Rightarrow 1-K/n \searrow \quad RSS_K/(1-\frac{K}{n})^2 \nearrow \quad$ GCV

* realistic estimate of expected squared prediction error

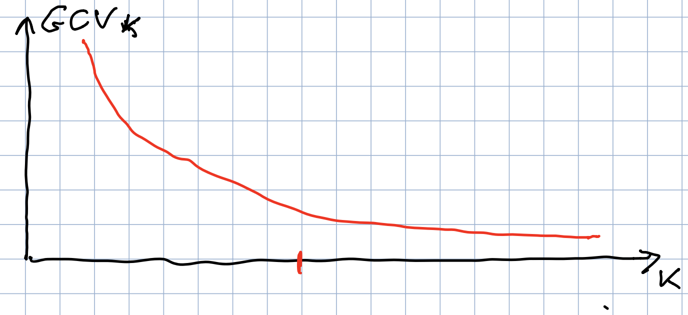
Problem: When used in conjunction with variable selection (like stepwise or all subsets) is biased estimate of expected squared prediction error.

Cross-validation assumes that $\hat{y} = Hy$ and H does NOT depend on y (the responses for the training sample.

This is not true for Turbo: For given K the colums of X (and therefor H) are chosen by using y.

When applied in conjuction with Turbo

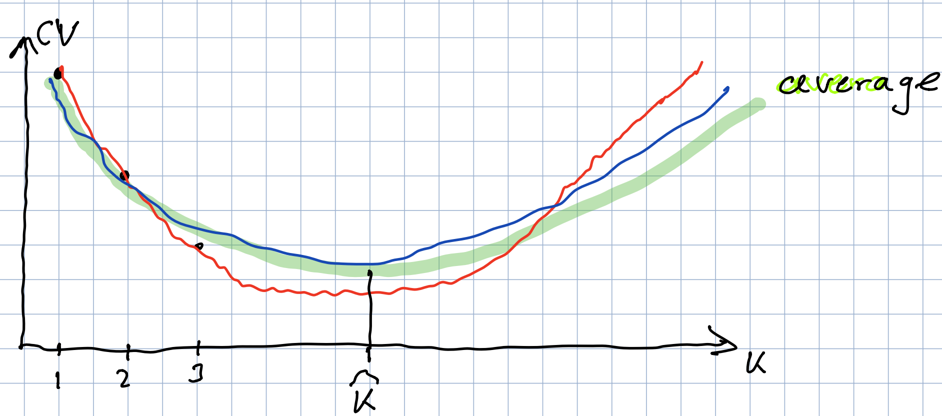
GCV_K may be monotonically decreasing in K



Fair CV:

Ramdomly split TS into m equally sized subsets $S_1 \dots S_m$

- Set S_1 aside
- Run stepwise selection of $S / S_1 \Rightarrow$ get sequence of models
- Use all these models to predict S_1



Once we have $\hat{K}$, we run forward selection on all the training data till we have $\hat{K}$ basis functions in model

Moving beyond 1 predictor

Turbo can easily construct additive prediction rules

$$f(\underline{x}) = f_1(x_1) + \underbrace{f_2(x_2) + f_m(x_m)} \Leftarrow$$

where m is the number of predictor variables

choose dichonory

$$B_1^1(x_1) \dots B_{p_1}^1(x_1) \leftarrow \text{depend on } x_1$$

$$B_1^2(x_2) \dots B_{p_2}^2(x_2)$$

$$\vdots$$

$$B_1^m(x_m) \quad B_{p_m}^m(x_m) \leftarrow \text{depend on } x_m$$

Total of $p_1 + p_2 + \dots p_m$ basis functions

TS $(\underline{x}_1, y_1) \dots (x_n, y_n)$ $\underline{x}_i \in \mathbb{R}^m$