Projection Pursuit Regression

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Additive forms in functions of derived variables

For $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M$ unit p-vectors of parameters, we might consider as predictors fitted versions of the form

$$f(\mathbf{x}) = \sum_{m=1}^{M} g_m(\mathbf{w}'_m \mathbf{x}) \quad (1)$$

This is an additive form in the derived variables $v_m = \mathbf{w}'_m \mathbf{x}$. The functions g_m and the directions $\mathbf{w}_m$ are to be fit from the training data. The $M = 1$ case of this form is the "single index model" of econometrics.

Fitting for *M=1*

How does one fit a predictor of form (1)? Consider first the $M = 1$ case. Given $\mathbf{w}$, we simply have pairs (v_i, y_i) for $v_i = \mathbf{w}'\mathbf{x}_i$ and a 1-dimensional smoothing method can be used to estimate g. On the other hand, given g, we might seek to optimize $\mathbf{w}$ via an iterative search. A Gauss-Newton algorithm can be based on the first order Taylor approximation

$$g(\mathbf{w}'\mathbf{x}_i) \approx g(\mathbf{w}'_{\text{old}}\mathbf{x}_i) + g'(\mathbf{w}'_{\text{old}}\mathbf{x}_i)(\mathbf{w}' - \mathbf{w}'_{\text{old}})\mathbf{x}_i$$

so that

$$\sum_{i=1}^{N} (y_i - g(\mathbf{w}'\mathbf{x}_i))^2$$
$$\approx \sum_{i=1}^{N} (g'(\mathbf{w}'_{\text{old}}\mathbf{x}_i))^2 \left(\left(\mathbf{w}'_{\text{old}}\mathbf{x}_i + \frac{y_i - g(\mathbf{w}'_{\text{old}}\mathbf{x}_i)}{g'(\mathbf{w}'_{\text{old}}\mathbf{x}_i)} \right) - \mathbf{w}'\mathbf{x}_i \right)^2$$

Fitting for *M=1* cont.

We may then update $\mathbf{w}_{\text{old}}$ to $\mathbf{w}$ using the closed form for weighted (by $\left(g'\left(\mathbf{w}'_{\text{old}}\mathbf{x}_i\right)\right)^2$) no-intercept regression of

$$\left(\mathbf{w}'_{\text{old}}\mathbf{x}_i + \frac{y_i - g\left(\mathbf{w}'_{\text{old}}\mathbf{x}_i\right)}{g'\left(\mathbf{w}'_{\text{old}}\mathbf{x}_i\right)}\right)$$

on $\mathbf{x}_i$. (Presumably one must normalize the updated $\mathbf{w}$ in order to preserve unit length property of the $\mathbf{w}$ in order to maintain a stable scaling in the fitting.) The g and $\mathbf{w}$ steps are iterated until convergence. Note that in the case where cubic smoothing spline smoothing is used in projection pursuit, g' will be evaluated as some explicit quadratic and in the case of locally weighted linear smoothing,

$$\hat{f}_\lambda\left(x_0\right) = \left(1, x_0\right)\left(\mathbf{B}'\mathbf{W}\left(x_0\right)\mathbf{B}\right)^{-1}\mathbf{B}'\mathbf{W}\left(x_0\right)\mathbf{Y}$$

will need to be differentiated in order to evaluate the derivative g'.

Fitting for *M>1*

When $M > 1$, terms $g_m(\mathbf{w}_m'\mathbf{x})$ are added to a sum of such in a forward stage-wise fashion. HTF provide some discussion of details like readjusting previous g's (and perhaps $\mathbf{w}$'s) upon adding $g_m(\mathbf{w}_m'\mathbf{x})$ to a fit, and the choice of M.