Projection Pursuit Regression

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- Projection pursuit regression (PPR) was introduced by Friedman (1981).
- Despite the fact that it can be very effective, and has been around since 1981, most of you probably have not heard about PPR.
- The idea is quite straightforward.
- However, there are implementation subtleties and interpretation is generally challenging.

The Idea

- We want to predict Y based on p-dimensional $\mathbf{X}$.
- Using notation similar to Hastie et al. (2009)[a], let $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_M$ be (unknown, p-dimensional) unit vectors.
- The PPR model is given by

$$f(\mathbf{X}) = \sum_{m1}^{M} g_m(\boldsymbol{\omega}'_m \mathbf{X}) = \sum_{m1}^{M} g_m(V_m),$$

where $V_m = \boldsymbol{\omega}'_m \mathbf{X}$ and $g_1(), \ldots, g_m()$ are unspecified functions.

[a] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer: NY.

The Unspecified Functions $g_m()$

- $g_m(V_m)$ is called a ridge function, and it needs to be estimated.
- $V_m = \boldsymbol{\omega}'_m \mathbf{X}$ is the projection of $\mathbf{X}$ onto $\boldsymbol{\omega}_m$.
- We want to find "good" $\boldsymbol{\omega}_m$.
- Hence the name, projection pursuit.
- Let's look at two examples of ridge functions, taken from Hastie et al. (2009).

Examples of Ridge Functions

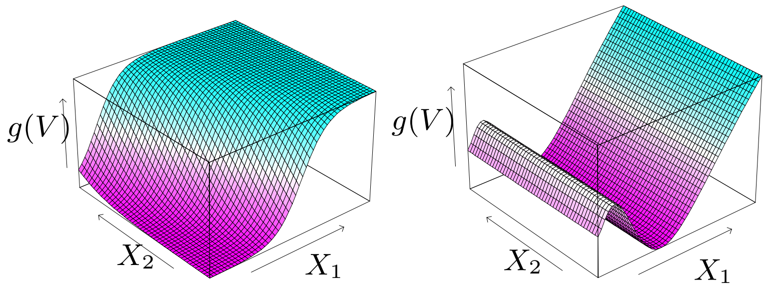

FIGURE 11.1. *Perspective plots of two ridge functions.*
(Left:) $g(V) = 1/[1 + \exp(-5(V - 0.5))]$, *where* $V = (X_1 + X_2)/\sqrt{2}$.
(Right:) $g(V) = (V + 0.1)\sin(1/(V/3 + 0.1))$, *where* $V = X_1$.

Model Fitting

- Suppose we observe $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$.

- The idea is to fit the PPR model by choosing $g_1(), \ldots, g_M()$ and $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_M$ to make the error

$$\sum_{i=1}^{n} [y_i - f(\mathbf{x}_i)] = \sum_{i=1}^{n} \left[y_i - \sum_{m=1}^{M} g_m(\boldsymbol{\omega}'_m \mathbf{x}_i) \right]$$

  small.

- Note that M also needs to be chosen.

- See Hastie et al. (2009) for further details on model fitting in PPR.

Comments

- Note that for "large" M and sensible choice of $g_1(), \ldots, g_m()$, PPR can do a "good" job at approximating any continuous function in $\mathbb{R}^p$.
- In this sense, PPR can be viewed as a universal approximator.
- However, as mentioned at the outset of the class, interpretation is generally (very) difficult.
- If the goal is prediction — and modelling *per se* is not considered important — difficulties in interpretation may not matter.

Comments contd.

- Broadly, PPR is a non-linear statistical model.
- There is a relationship between PPR and neural networks...
- ...and we will look at neural networks next.
- But first, let's look at some PPR examples in R.