STAT40810 — Stochastic Models

Brendan Murphy

Week 5

Projection Pursuit Regression

Projection Pursuit Regression

- Projection pursuit regression (PPR) extends the multiple linear model in a different way.
- Instead of using a sum of smooth functions of each covariate separately, like in the GAM,

$$Y_i = \alpha + g_1(X_{i1}) + g_2(X_{i2}) + \cdots + g_K(X_{iK}) + \epsilon_i,$$

  where Y_i is the response variable and the covariates are $X_i = (X_{i1}, X_{i2}, \ldots, X_{iK})$.
- PPR writes the response variable in terms of smooth functions of linear combinations of variables.
- A PPR regression with one term has the form

$$Y_i = \alpha + f(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK}) + \epsilon_i,$$

  where $f(\cdot)$ is a smooth function.

Projection Pursuit Regression

- More generally PPR can have more than one term

$$Y_i = \alpha + \sum_{k=1}^{K} f_k(\beta_k' x_i) + \epsilon_i.$$

- That is, the response variable is a combination of smooth functions of projections of the data.

R Code

```
# Load the rock data
data(rock)

# Produce a pairs plot of the data
pairs(rock)

# Add a column with log(perm)
rock$lperm <- log(rock$perm)

# Fit a PPR model
fit <- ppr(lperm~area+peri+shape,data=rock,nterms=2)

# Examine the fit
summary(fit)

# Plot the smooth functions used in each term
par(mfrow=c(1,2))
plot(fit)
par(mfrow=c(1,1))
```

Projection Pursuit Regression: Results

```
Goodness of fit:
2 terms  3 terms  4 terms  5 terms
8.284056 5.506386 4.979411 4.491782

Projection direction vectors:
      term 1        term 2
area   5.957456e-04  5.488554e-05
peri  -1.602357e-03 -9.178419e-05
shape  9.999985e-01  1.000000e+00

Coefficients of ridge terms:
   term 1    term 2
1.6273943 0.5506679
```

Projection Pursuit Regression: Plot