Lecture 8

Beyond Linearity

ISLR 7, ESL 5,6,9

Jilles Vreeken
Krikamol Muandet

UNIVERSITÄT DES SAARLANDES
CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Polynomial Regression

Standard linear model

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

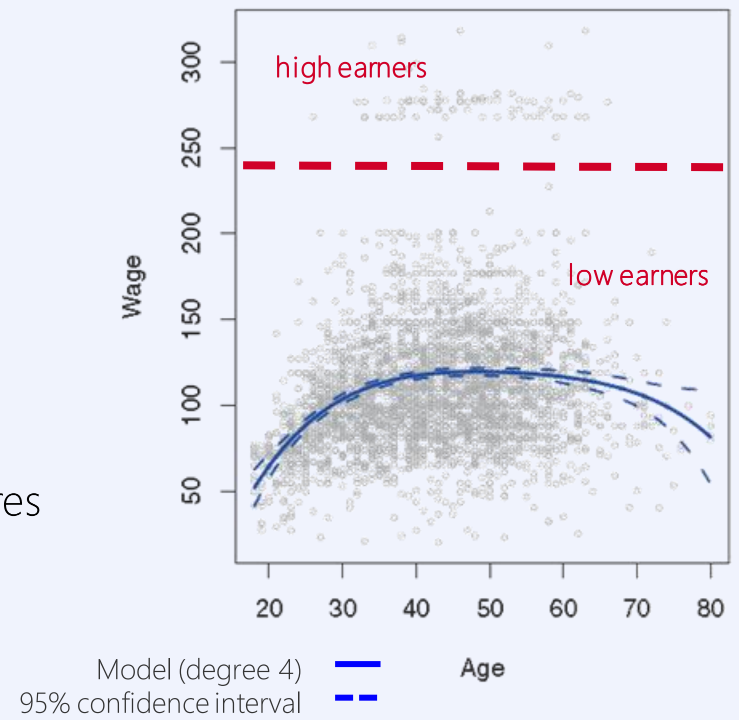
Polynomial regression

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \epsilon$

Model is still linear in the coefficients β_i!

- compute confidence bounds as before using pointwise variance from least squares

example regression on wage data



Step Functions

We convert a **continuous** to an **ordered categorical** variable (ordinal)

- create cutpoints $c_1, c_2, \ldots, c_K$ in the range of X
- construct $K+1$ new variables

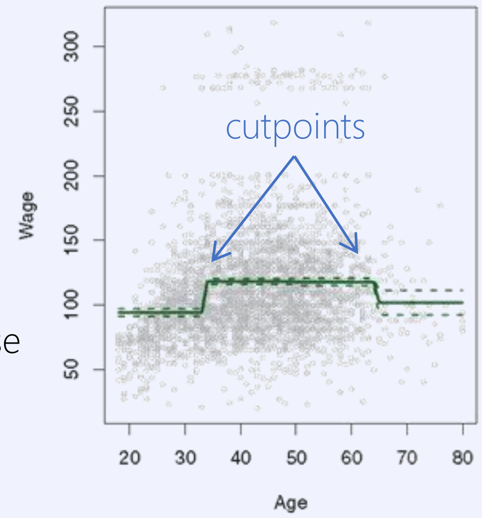
dummy variables

$$C_0(X) = I(X < c_1),$$
$$C_1(X) = I(c_1 \le X < c_2),$$
$$\ldots$$
$$C_{K-1}(X) = I(c_{K-1} \le X < c_K),$$
$$C_K(X) = I(c_K \le X)$$

$$\sum_{i=0}^{K} C_i = 1$$

- $I(\cdot)$ is the indicator function: 1 if its argument is true and zero otherwise

wage regression



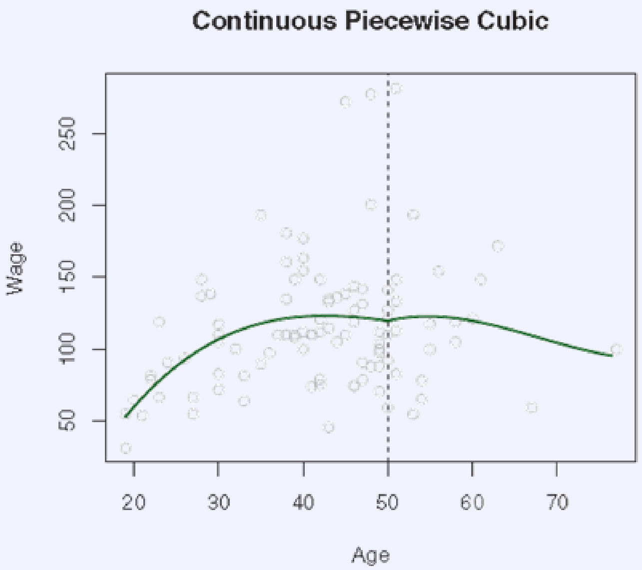
Regression $y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \epsilon_i$

- β_0 is the average of Y for all $X < c_1$
- β_j is the average increase in Y over β_0 for $c_j < X < c_{j+1}$

Regression Splines

Instead of fitting one high-degree polynomial,
we fit a low-degree polynomial ***per region*** of X

- make sure that the model is smooth at region boundaries
- that is, continuous and $d-1$ times continuously differentiable, where d is the degree of the polynomial

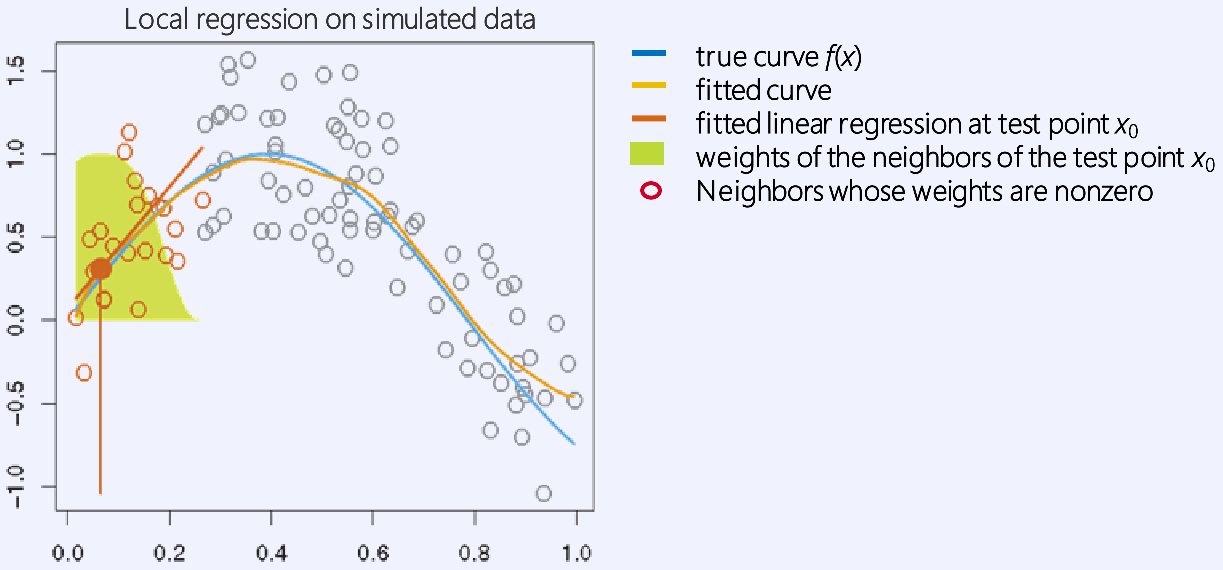


single cutpoint at ***age=50***

Local Regression

Extension of k-nearest neighbors

- fits not constant, but polynomial models based on the nearest neighbors of a test point
- weighs the contribution of neighbors by their distance to test point



(ISLR 7.6)

Generalized Additive Models (GAM)

General framework for including nonlinear basis functions into linear multivariate models

- generalize the linear model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

to

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \epsilon_i$$

All methods we discussed so far can be plugged into this scheme (!)

Generalized Additive Models (GAM)

Pros and cons of GAMs

- ➕ nonparametric, no need of trying out different model assumptions
- ➕ nonparametric, can afford more accurate predictions
- ➕ since model is additive, we can assess the influence of a variable while holding the other variables fixed
- ➕ smoothness of function f_j for variable X_j can be summarized via degrees of freedom
- ➖ restriction of the model to be additive, this can miss important interactions
  - but, we can add predictors like $X_j \times X_k$ fitted with e.g. two-dimensional splines

Generalized Additive Models (GAM)

GAMs for classification

- use logistic regression
- linear model $\log\left(\frac{p(x)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$
- generalized additive model $\log\left(\frac{p(x)}{1-p(X)}\right) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p)$

Example on the wage data: $p(X) = \Pr(\texttt{wage} > 250 \mid \texttt{year}, \texttt{age}, \texttt{education})$

- the GAM takes the form

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + f_1(\texttt{year}) + f_2(\texttt{age}) + f_3(\texttt{education})$$