Nonlinear Models

非线性模型

Outline

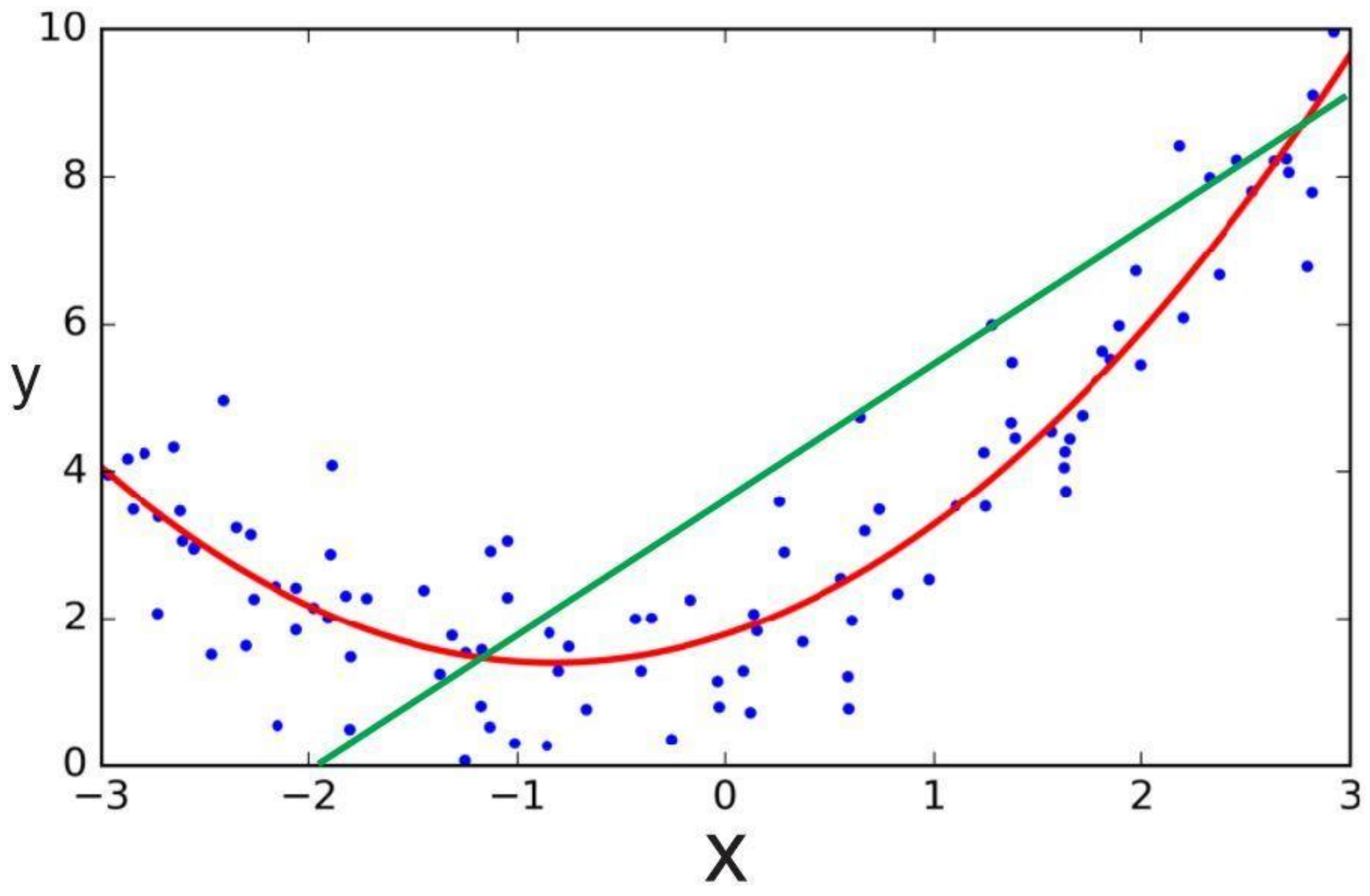
- 回顾：线性回归
- 多项式回归
- 分段函数拟合
- 基函数的形式
- 平滑样条
- 可加模型

回顾：线性回归

- 考虑随机变量(x, y), 其中(x, y) ∈ R.
- 在线性模型下，我们常假设

$$E[y|x] = \beta_0 + x\beta$$

- 针对复杂的数据，仅使用一条直线去拟合往往会导致较差的精度



- 由上图可清楚地看出，线性模型并不能够准确地刻画数据的变化趋势
- 这时我们需要考虑非线性模型来提高精度

多项式回归

- 考虑随机变量(x, y), 其中$x \in R^p, y \in R$.
- 多项式回归模型通过将线性回归函数替换为多项式回归函数来扩展简单线性回归，其所考虑的模型如下

$$y = \beta_0 + \sum_{j=1}^{d} \beta_j^T \mathrm{x}^{\mathrm{j}} + \epsilon$$

- 由上式可知，多元线性回归模型可以看作是多项式回归模型的一 个特例（ $d = 1$ ）。
- 因此可以采用与线性回归模型相同的估计方法得到有关回归系数的估计。

多项式回归

- 令 $\theta = (\beta_0, \beta_1^T, \dots, \beta_d^T)^T \in R^{dp+1}$ 以及 $\tilde{x} = (1, (x)^T, (x^1)^T, \dots, \left(x^d\right)^T)^T \in R^{dp+1}$,

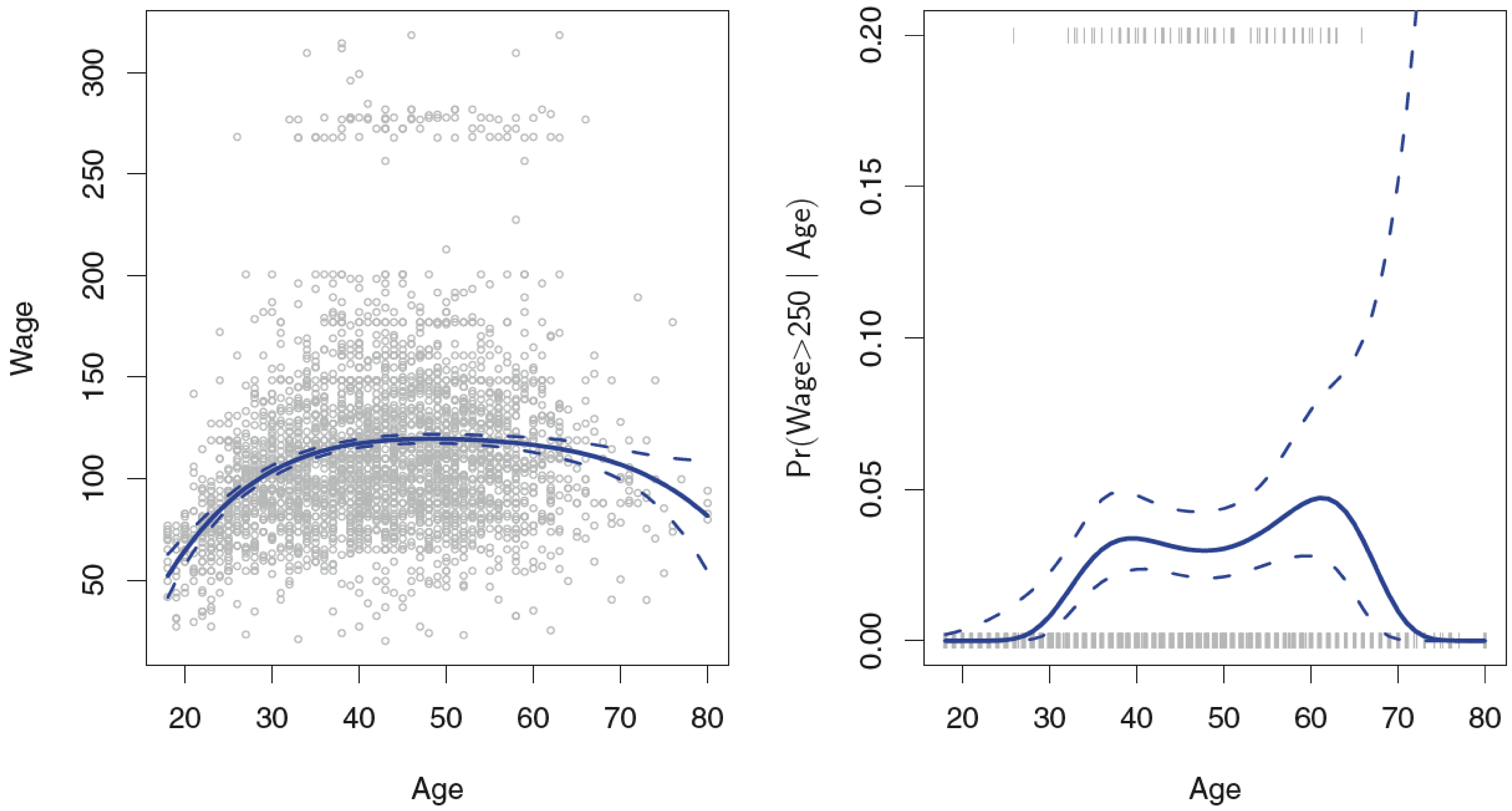
$$y = \beta_0 + \sum_{j=1}^{d} \beta_j^T \mathrm{x}^{\mathrm{j}} + \epsilon = \theta^T \tilde{x} + \epsilon.$$

- 给定样本 $Z^n = \{(y_i, x_i)\}_{i=1}^n$，类似于最小二乘估计方法，求解如下的优化问题

$$\sum_{i=1}^{n} (y_i - \theta^T \tilde{x}_i)^2$$

- 一般情况下，d 的取值为 2,3,4, 并且只有在极少情况下 $d \geq 4$。
- 通常，会预先将样本中心化。

一个采用四阶多项式逻辑回归的例子



- 左上图表明数据中存在两个不同的组，即工资可能来自两组不同的人群。
- 因而采用回归模型拟合上述数据并不是一个好的选择。

多项式逻辑回归

- 构造一个二分类响应变量：

$y=1$： 表示年收入超过 25 万元；

$y=0$： 表示年收入低于 25 万元。

- 给定样本 $Z^n=\{(y_i,x_i)\}_{i=1}^n$ ， 多项式逻辑回归考虑如下模型

$$\mathrm{logit}(\Pr(y_i=1|x_i))=\beta_0+\sum_{j=1}^{d}\beta_j^T x_i^j=\theta^T\widetilde{x}_i$$

- 使用极大似然估计得到相关参数的估计值

$$\min_{\theta\in R^{dp+1}}-\frac{1}{n}\sum_{i=1}^{n}\{y_i\widetilde{x}_i{}^T\theta-\log(1+e^{\widetilde{x}_i{}^T\theta})\}$$

- 可使用NR算法或梯度下降法。

交叉项回归模型

- 上面的模型中，只考虑了每个自变量单独对因变量的影响；
- 在实际数据处理中，如在基因数据中，基因之间往往有着交互的作用。
- 当模型中有多个预测变量时，可以考虑变量之间的交互作用。
- 比如，假设模型中有两个自变量 x_1, x_2，可以考虑将它们之间的交互项 (interaction term) 引入模型中，即
$$E[y|x] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma_{12} x_1 x_2$$
- 令$\theta = (\beta_0, \beta_1, \beta_2, \gamma_{12})^T, \tilde{x} = (1, x_1, x_2, x_1 x_2)^T$，可得到 $E[y|x] = \theta^T \tilde{x}$

交叉项回归模型

- 给定训练样本 $Z^n = \{(y_i, x_i)\}_{i=1}^n$，其中 $x_i = (x_{i1}, x_{i2}) \in R^2$.
- 可以参考线性回归的方法得到有关系数的估计

$$\min_{\theta \in R^4} \sum_{i=1}^{n} (y_i - \theta^T \tilde{x}_i)^2$$

其中$\tilde{x}_i = (1, x_{i1}, x_{i2}, x_{i1}x_{i2})^T$。

分段常函数

- 分段常函数提供了一种可以局部近似非线性结构的有效方法
- 其将连续型变量转化为有序的二进制变量
- 考虑随机变量(y, x)，其中$x, y \in R$
  - 令 $c_1, \ldots, c_K$表示 x 取值范围内的 K 个截断点
  - 构造 $K+1$ 个局部函数

$$
\begin{aligned}
C_0(x) &= I(x < c_1) \\
C_1(x) &= I(c_1 \leq x < c_2) \\
&\vdots \\
C_K(x) &= I(x \geq c_K),
\end{aligned}
$$

（其中$I(\cdot)$表示示性函数）

分段常函数

- 显然，对于任意 x 的一个观测值x_i，其只能属于这 $K+1$ 个区间中的一个，即

$$\mathrm{C_0(x_i) + C_1(x_i) + C_2(x_i) + \cdots + C_K(x_i) = 1}$$

- 因而，假设数据由如下模式产生

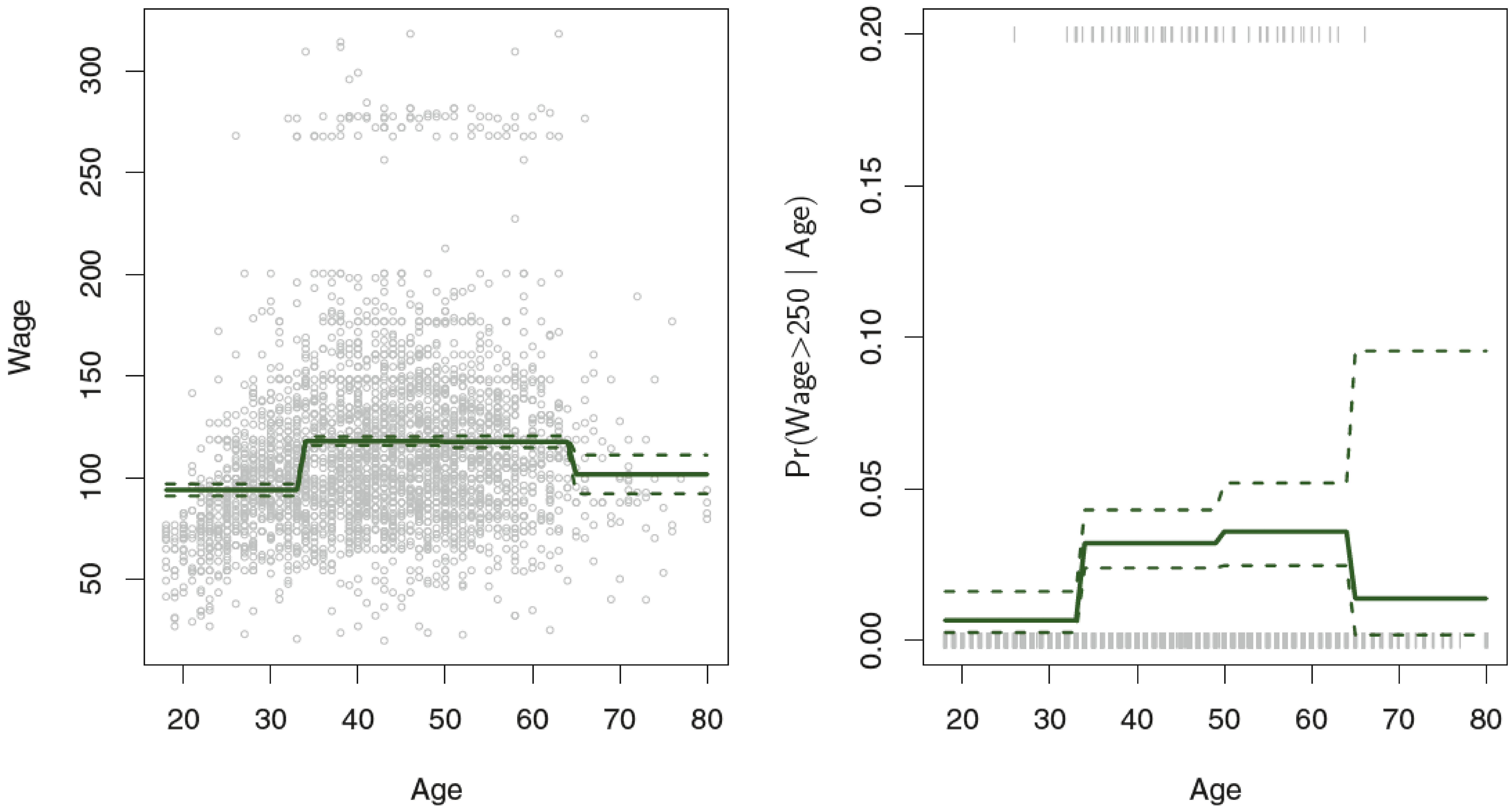
$$\mathrm{y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \epsilon_i}$$

- 拟合函数在点 x_0处的估计值为

$$\widehat{\mathrm{f}}(\mathrm{x_0}) = \begin{cases} \widehat{\beta}_0 + \widehat{\beta}_\mathrm{k}, & \text{如果 } \mathrm{c_k \le x_0 < c_{k+1}}, \\ \widehat{\beta}_0, & \text{如果 } \mathrm{x_0 < c_1}. \end{cases}$$

- 显然，回归函数由许多个分段常数函数所组成
- 有关参数的估计即为落在此区间内样本响应变量的均值

使用分段常函数函数拟合的例子



分段多项式

- 分段多项式模型在 x 的定义域内分段拟合一些低阶的多项式
- 考虑一维情况 $x \in R$，将 x 的定义域分为两部分，并且在每段内拟合一个三次多项式回归模型
- 因而，假设数据由如下模型产生

$$
\begin{aligned}
y_i = \beta_{01} + \beta_{11} x_i + \beta_{21} x_i^2 + \beta_{31} x_i^3 + \beta_{02} C(x_i) \\
+ \beta_{12} C(x_i) x_i + \beta_{22} C(x_i) x_i^2 + \beta_{32} C(x_i) x_i^3 + \epsilon_i,
\end{aligned}
$$

其中$C(x_i) = I(x_i \geq c)$，以及 c 表示进行分段的节点（Knot）

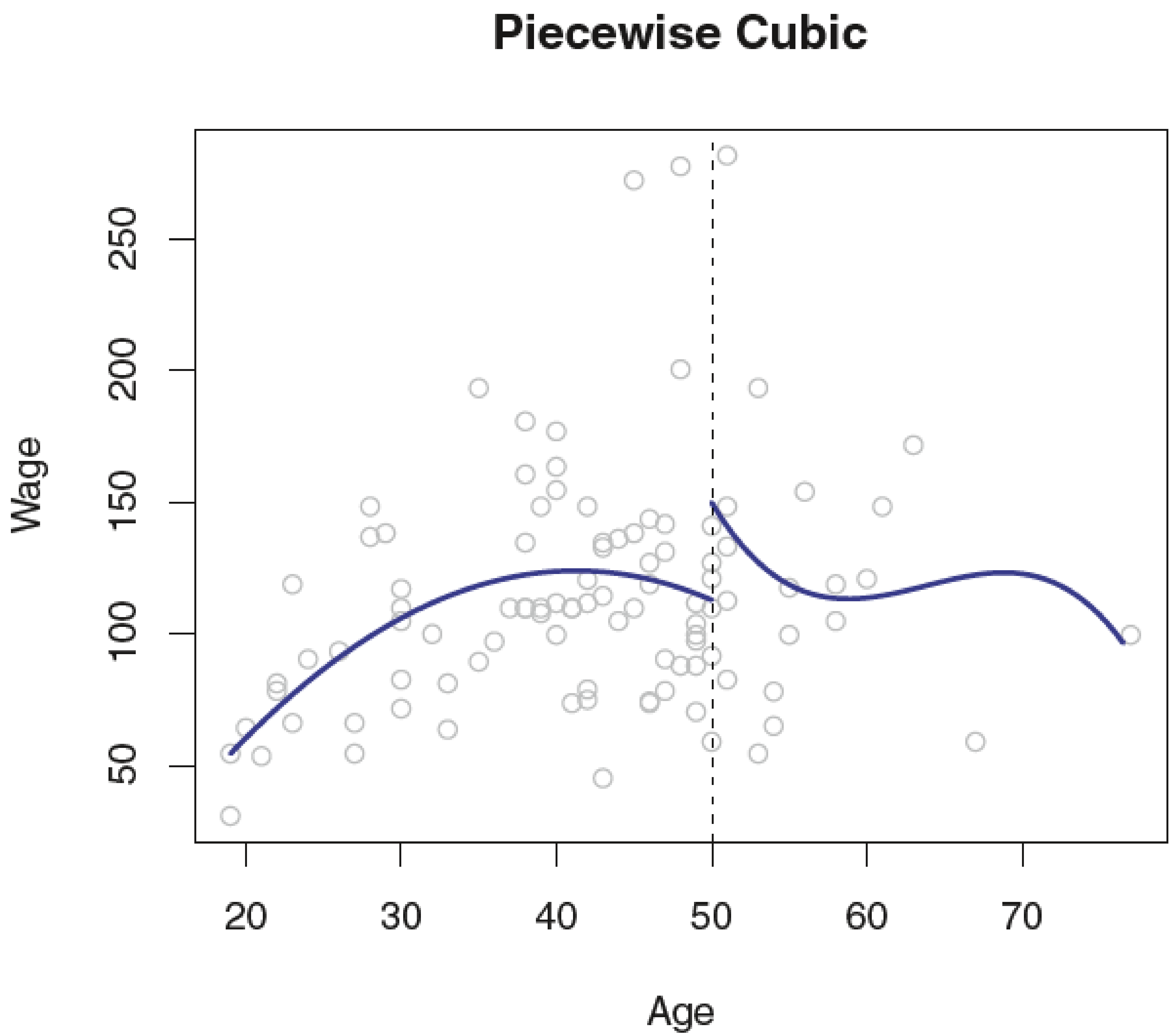
- 上述模型可写为

$$
f(x_i) = \begin{cases} \beta_{01} + \beta_{11} x_i + \beta_{21} x_i^2 + \beta_{31} x_i^3 & \text{如果 } x_i < c \\ \beta_{01} + \beta_{02} + (\beta_{11} + \beta_{12}) x_i + (\beta_{21} + \beta_{22}) x_i^2 + (\beta_{31} + \beta_{32}) x_i^3 & \text{如果 } x_i \geq c. \end{cases}
$$

- 采用更多的节点会产生一个更加灵活的（flexible）拟合函数

一个二分段多项式拟合的例子

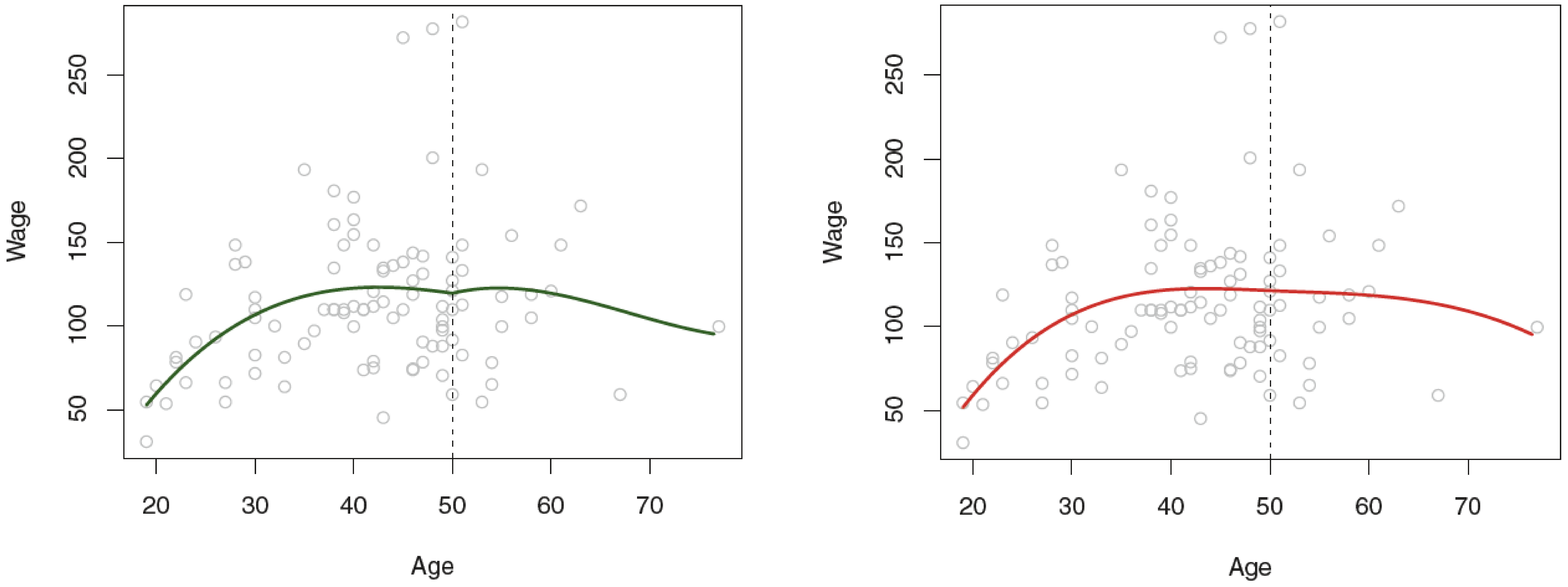
- 下图展示了使用两段三次多项式拟合年龄与收入的关系。



连续分段多项式

- 上图中拟合曲线存在一些问题：
  - 预测的工资收入在年龄 50 时有一个不连续的跳跃
- 为了解决上述问题，可以针对所拟合的多项式函数添加相关的约束条件
- 一个常常用到的约束条件为：拟合曲线在节点处必须是连续的
- 此外，为了使所得到的拟合函数更加平滑，还可以要求拟合的分段多项式函数的导数必须是连续的，这样就得到了非线性回归中常用的平滑样条函数

两个例子



左图：连续的分段多项式函数；右图：有着连续的一阶以及二阶导数的分段多项式函数

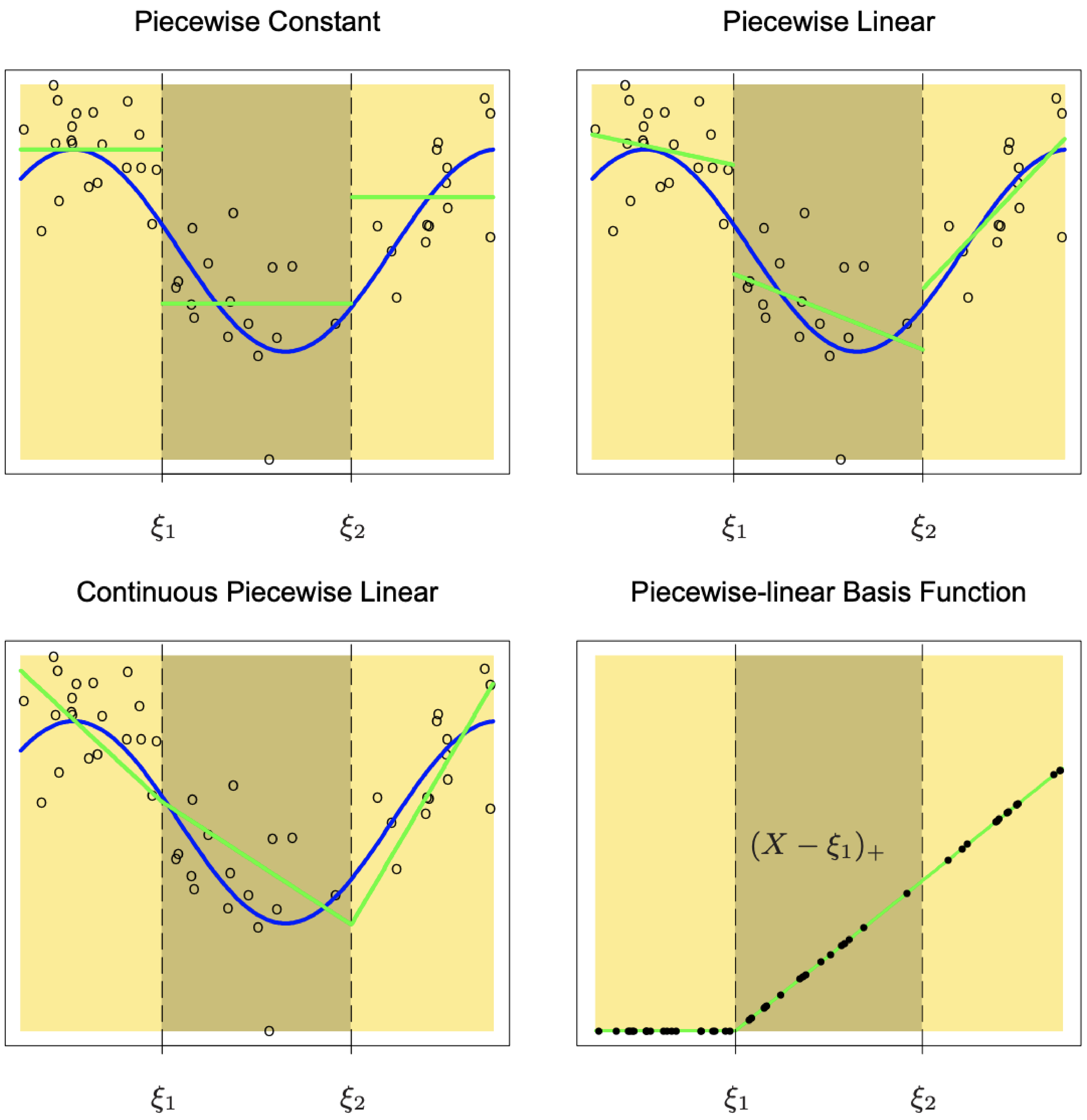
基函数

- 以上分段常函数以及分段多项式函数的例子中，都是在 x 定义域内的每一分段内拟合一个单独的函数
- 更一般化来看，可以将这些拟合的函数看作一组基函数
- 通常，假设如下的一般形式：

$$y_i = f^\star(x_i) + \epsilon_i,\ \text{其中}\ f^\star(x_i) = \sum_{k=1}^{m} \beta_k h_k(x_i),$$

- 这里$h_1(\cdot), \ldots, h_m(\cdot)$为固定且已知的基函数
  - 分段常数函数：$h_k(x) = I(c_k \le x < c_{k+1})$;
  - 分段多项式函数： $h_k(x) = x^{\mathrm{k}}$;
  - 平滑样条函数：三次样条，B样条等。

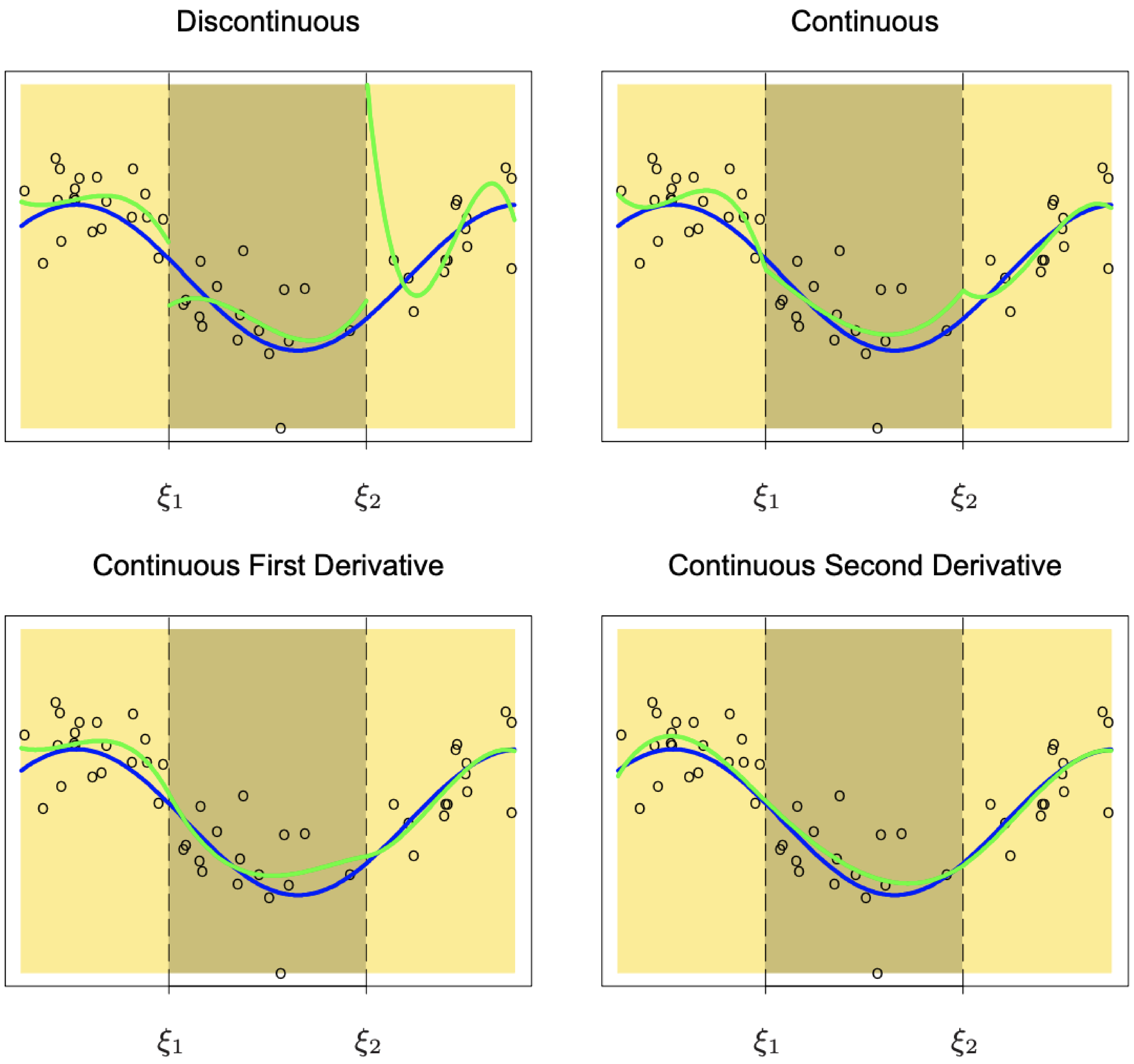
例子：分段线性拟合


Piecewise Constant
ξ_1
ξ_2
Piecewise Linear
ξ_1
ξ_2
Continuous Piecewise Linear
ξ_1
ξ_2
Piecewise-linear Basis Function
$(X - \xi_1)_+$
ξ_1
ξ_2

例子：分段线性拟合

- 蓝线表示真实模型，误差项服从高斯分布，ξ_1, ξ_2表示事先选定的节点（knots）;
- 第一张图：$h_1(x) = I(x < \xi_1)$, $h_2(x) = I(\xi_1 \le x < \xi_2)$ 以及 $h_3(x) = I(\xi_2 \le x)$;
  - 有关 β 的估计即为落在区间内所有样本 y 的均值;
- 第二张图：除上述 $h_1(x), h_2(x), h_3(x)$ 之外，引入 $h_4(x) = h_1(x)x$, $h_5(x) = h_2(x)x$ 以及 $h_6(x) = h_3(x)x$;
- 第三张图：添加连续性条件在节点处值相等;
  - 等价于引入：$h_1(x) = 1, h_2(x) = x, h_3(x) = (x - \xi_1)_+$ 以及 $h_4(x) = (x - \xi_2)_+$;
- 在处理实际数据中，我们更倾向于得到一些更为平滑的拟合;

例子：平滑拟合



Discontinuous
Continuous
ξ_1
ξ_2
ξ_1
ξ_2
Continuous First Derivative
Continuous Second Derivative
ξ_1
ξ_2
ξ_1
ξ_2

例子：平滑拟合

- 蓝线表示真实模型，误差项服从高斯分布，ξ_1, ξ_2 表示事先选定的节点 (knots)；
- 第四张图中的函数也被成为三次样条 (Cubic spline)：

$$h_1(x) = 1,\ h_2(x) = x,\ h_3(x) = x^2$$
$$h_4(x) = x^3,\ h_5(x) = (x - \xi_1)_+^3,\ h_6(x) = (x - \xi_2)_+^3$$

- 上述基函数所构成的拟合函数具有连续的一阶及二阶导数， 也被成为三次样条函数

三次样条

- 三次样条（Cubic spline）是具有连续一阶和二阶导数的连续分段多项式
- 有着 K 个节点（knots）的三次样条函数可表达为

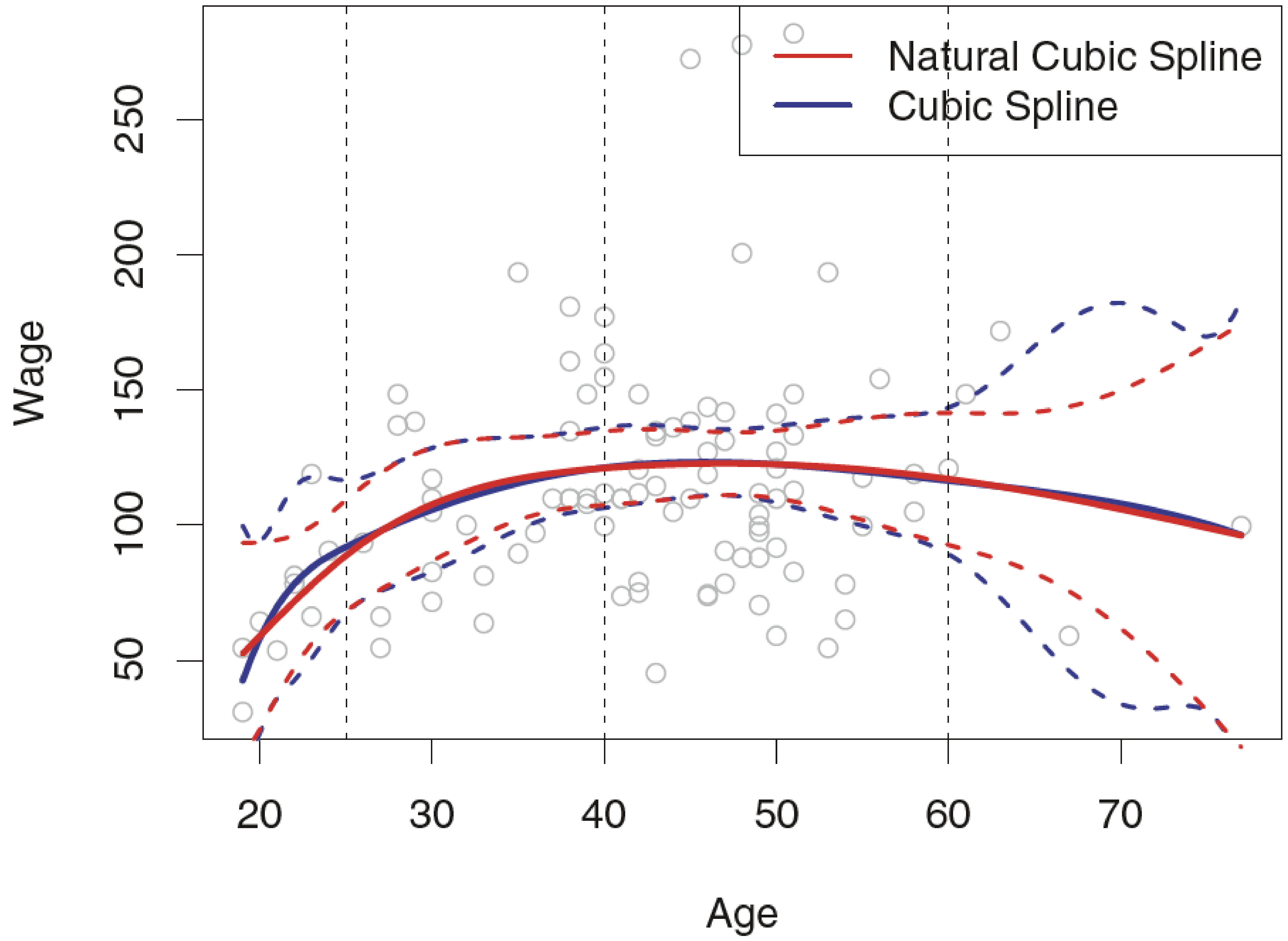
$$\mathrm{y_i} = \sum_{\mathrm{k}=1}^{\mathrm{K}+4} \beta_{\mathrm{k}} \mathrm{h_k}(\mathrm{x_i}) + \epsilon_{\mathrm{i}},$$

- 其中基函数为 $1, x, x^2, x^3$ 以及

$$\mathrm{h_k}(\mathrm{x}) = (\mathrm{x} - \xi_{\mathrm{k}-4})_+^3 = \max\left((\mathrm{x} - \xi_{\mathrm{k}-4})^3, 0\right), \mathrm{k} = 5, \ldots, \mathrm{K} + 4$$

- 这里 ξ_k 表示所选取的K个节点

三次样条

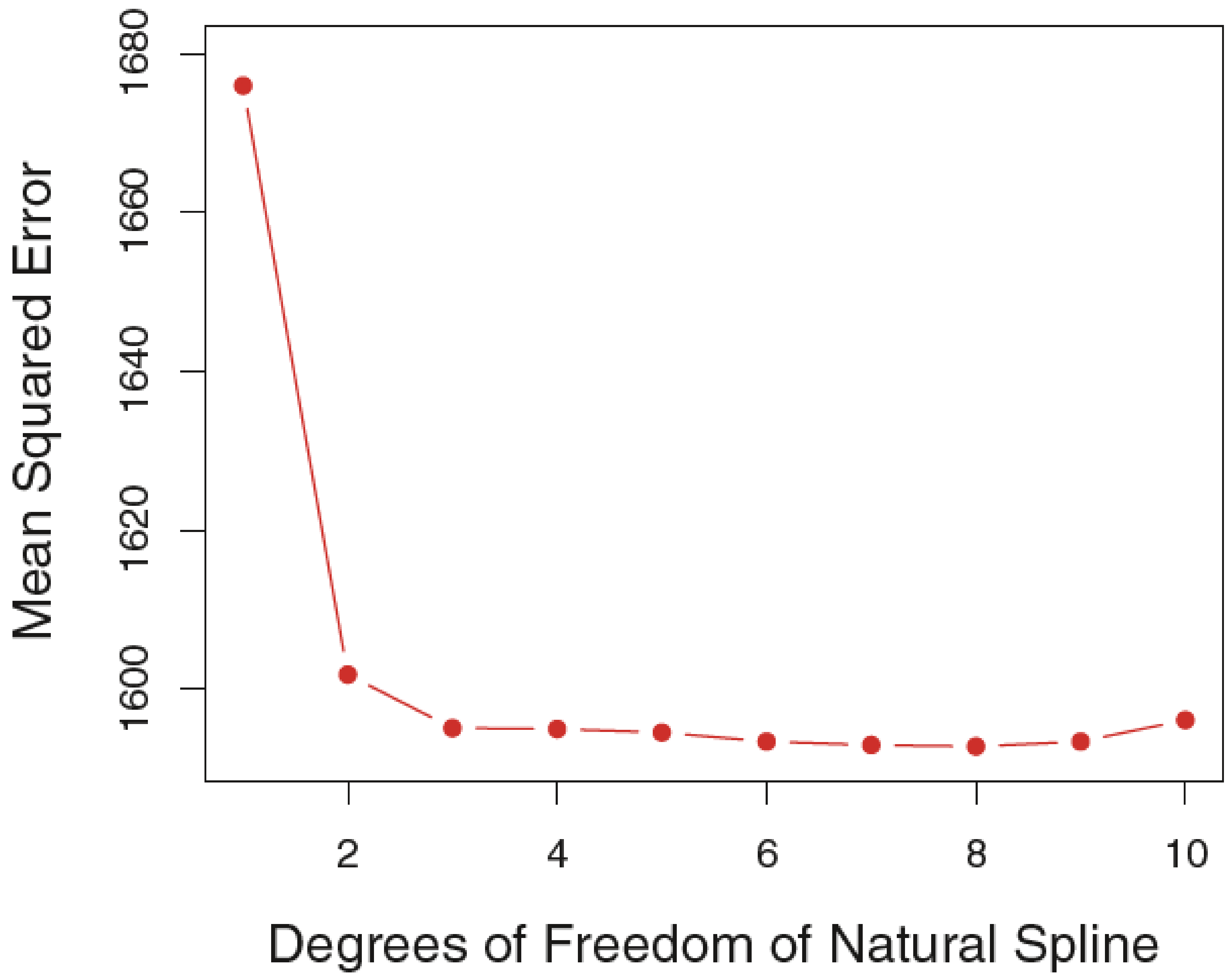


- 自然样条函数在预测变量取值的边界可能有着较大的方差
- 自然三次样条要求拟合模型在预测变量边界处是线性的

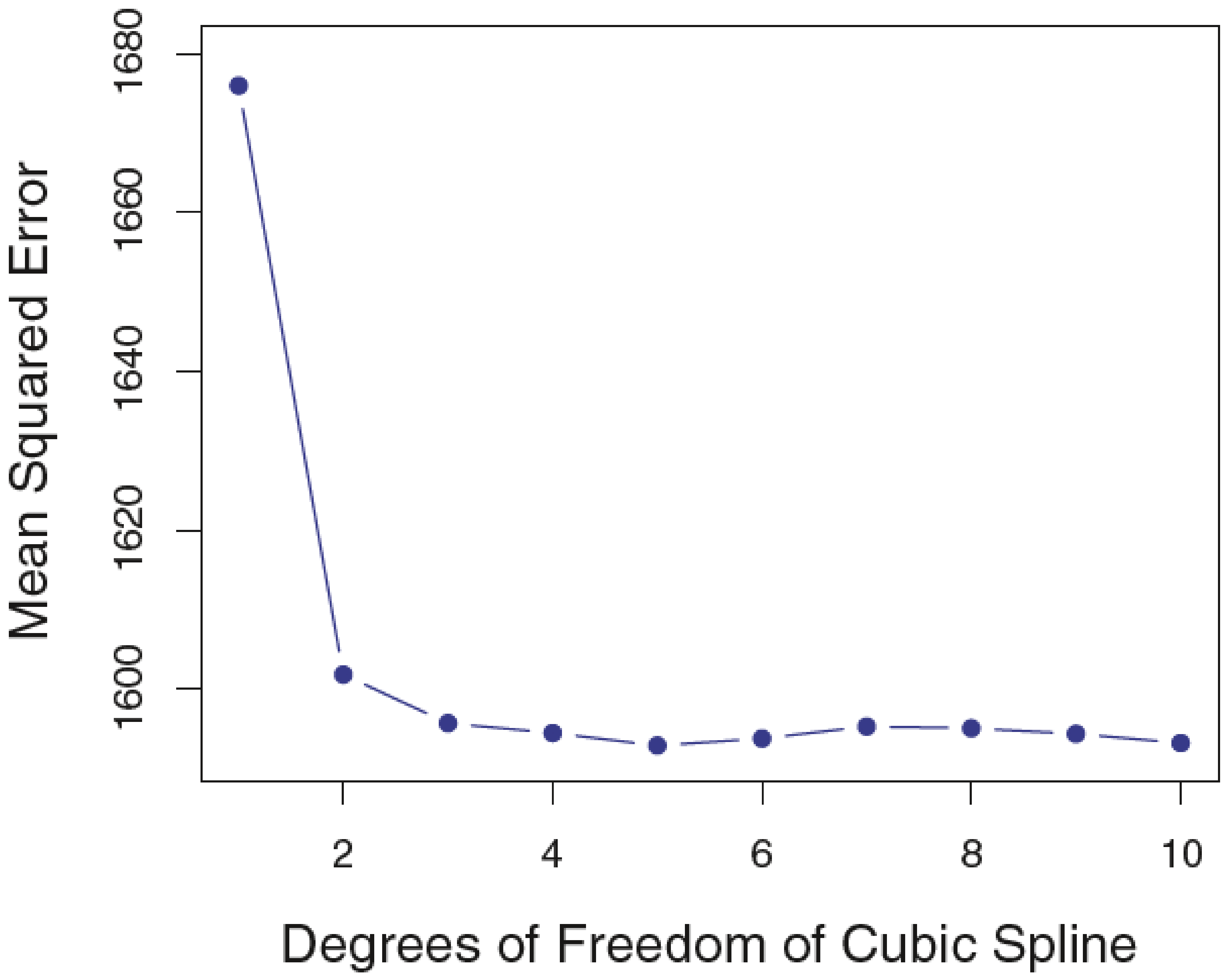
节点个数的选取

- 上述基于样条函数方法的拟合效果在一定程度上均依赖于节点个数的选取.
- 节点（knots)可以通过均匀设计的方式进行选取
- 比如针对上图中，如果 $k = 3$，可以选取有关变量 Age 的第25，50以及75分位点作为结点
- 也可以采取一些更为复杂的方法来设定结点以达到不同的目的
- 事实上，有关结点个数的设定会影响拟合模型的复杂度(自由度，因而可以通过交叉验证的方法选取最优的结点数)

节点个数的选取



Mean Squared Error
1600
1620
1640
1660
1680
2
4
6
8
10
Degrees of Freedom of Natural Spline



Mean Squared Error
1600
1620
1640
1660
1680
2
4
6
8
10
Degrees of Freedom of Cubic Spline

平滑样条

- 基于三次样条的非线性模型需要通过一些准则来得到最优的节点数，因而有着较大的计算量（如利用交叉验证等方法）
- 平滑样条通过采用最大节点数来避免有关节点选择的问题，并且通过引入正则化项来控制拟合模型的复杂度

平滑样条

- 平滑样条尝试在所有具有二阶导数连续的函数中找到一个函数 f 使其最小化如下问题

$$\min_{f} \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int (f''(t))^2 dt \tag{1}$$

- 第一项促使 f 能够尽可能好地拟合数据
- 第二项为正则化项，其惩罚有关 f 的复杂度。
  - 其中 $f''(t)$ 度量了有关 f 的平滑度（二阶可导）
- $\lambda \geq 0$为可调参数:
  - 当 $\lambda = 0, f$ 可以是任何在样本间的插值函数;
  - 当 $\lambda = \infty, f$ 只能是线性函数(二阶导数需全为零).

平滑样条

- 根据不同的数据类型，可以考虑使用不同的损失函数：
  - 如平方损失函数 、分位数损失函数 、逻辑损失函数 、hinge 损失函数
- 事实上，(1) 所考虑的函数空间为一个无穷维的函数空间（Soblev 函数空间）
- 更重要的是 (1) 有着唯一的，以 $x_1, \ldots, x_n$ 为节点的显示解，其可表达为

$$g(x) = \sum_{j=1}^{n} n_j(x)\theta_j, \qquad (2)$$

- 这里 $n_j(x) = n_{x_j}(x)$ 表示以 x_j 为节点的基底函数（自然样条函数的一组基）。

平滑样条

- 将 (2) 代入 (1) 中，可以得到

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} (\mathrm{y} - \mathrm{N}\boldsymbol{\theta})^{\mathrm{T}}(\mathrm{y} - \mathrm{N}\boldsymbol{\theta}) + \lambda\boldsymbol{\theta}^{\mathrm{T}}\Omega_{\mathrm{n}}\boldsymbol{\theta},$$

$$\text{其中 } \mathrm{N} \in \mathcal{R}^{\mathrm{n}\times\mathrm{n}},\ \{\mathrm{N}\}_{\mathrm{ij}} = \mathrm{n_j}(\mathrm{x_i}) \text{ 以及 } \{\Omega_{\mathrm{n}}\}_{\mathrm{jk}} = \int \mathrm{N}_{\mathrm{j}}''(\mathrm{t})\mathrm{N}_{\mathrm{k}}''(\mathrm{t})\mathrm{dt}.$$

- 求解可得

$$\widehat{\boldsymbol{\theta}} = (\mathrm{N}^{\mathrm{T}}\mathrm{N} + \lambda\Omega_{\mathrm{n}})^{-1}\mathrm{N}^{\mathrm{T}}\mathrm{y}.$$

- 针对新的样本 x_0，所得到的估计为

$$\widehat{\mathrm{f}}(\mathrm{x}_0) = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{n_j}(\mathrm{x}_0)\widehat{\boldsymbol{\theta}}_{\mathrm{j}}.$$

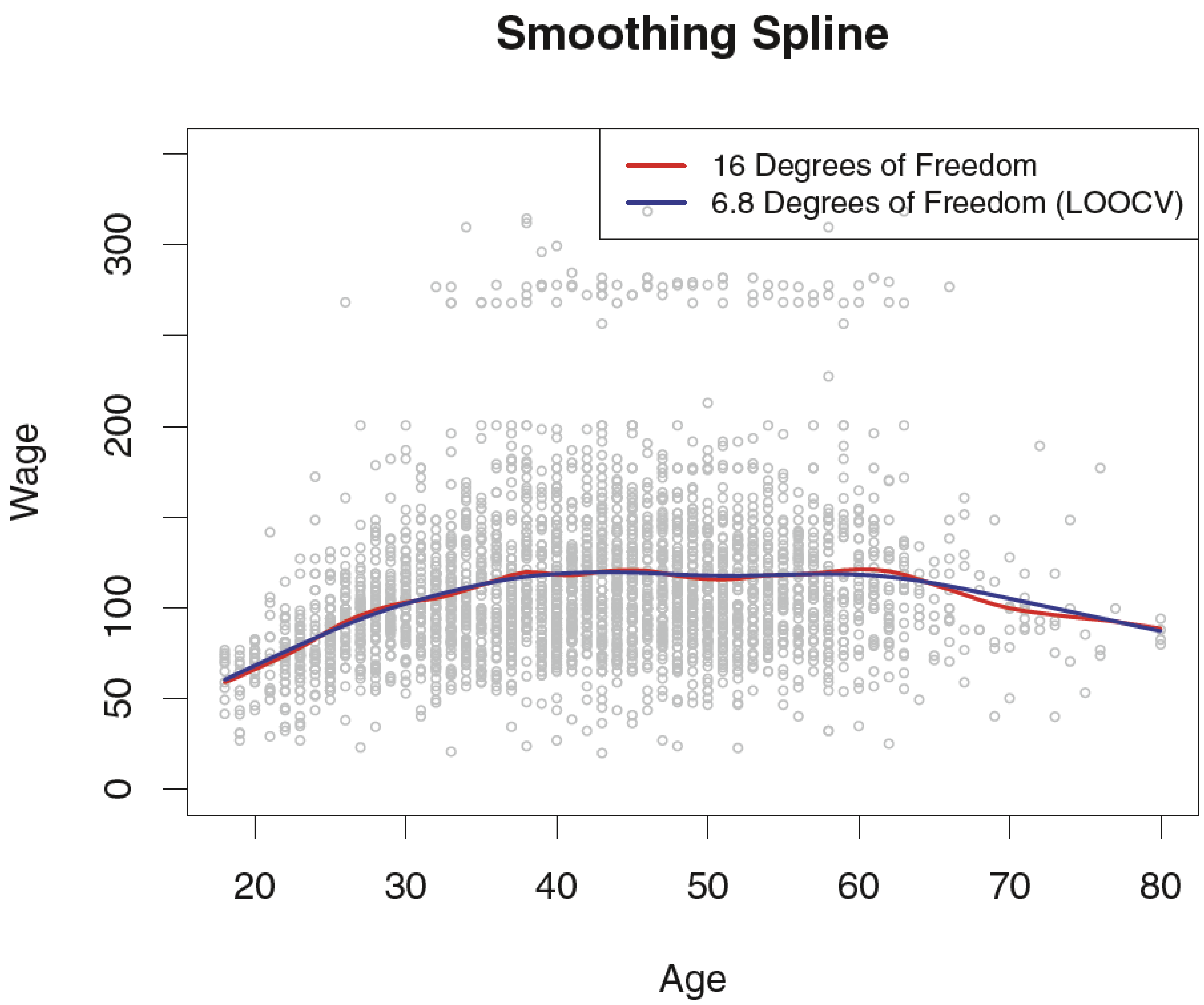
注意事项

- 上式中得到的 $\hat{f}(x)$可以被证明具有以下几点性质：
  - $\hat{f}(x)$ 为一个取 $x_1, \ldots, x_n$ 为节点的分段三次多项式函数；
  - 在每一个节点处，$\hat{f}(x)$都有着连续的一阶以及二阶导数；
  - 它在端点结点以外的区域里是线性的
- 平滑样条可以看作是一个以$x_1, \ldots, x_n$为节点的自然三次样条
  - 但是它与三次样条有着不同的基函数
- 事实上，其可以看作是一个具有收缩估计结果的自然三次样条， 其收缩水平由 λ 所控制

待调参数

- 当 $\lambda = 0$ 时，正则化项将不起作用，并且 f 将会在训练集中进行插值拟合.
- 当 $\lambda \to \infty$ 时，平滑样条退化为简单线性回归
- 可见，λ 控制着平滑样条的偏差与方差平衡，
  - 不同的 λ 值会导致不同的估计函数$\hat{f}_\lambda$
  - 最优的 λ 可以通过交叉验证进而得到

一个平滑样条的例子



Smoothing Spline
16 Degrees of Freedom
6.8 Degrees of Freedom (LOOCV)
Wage
0
50
100
200
300
20
30
40
50
60
70
80
Age

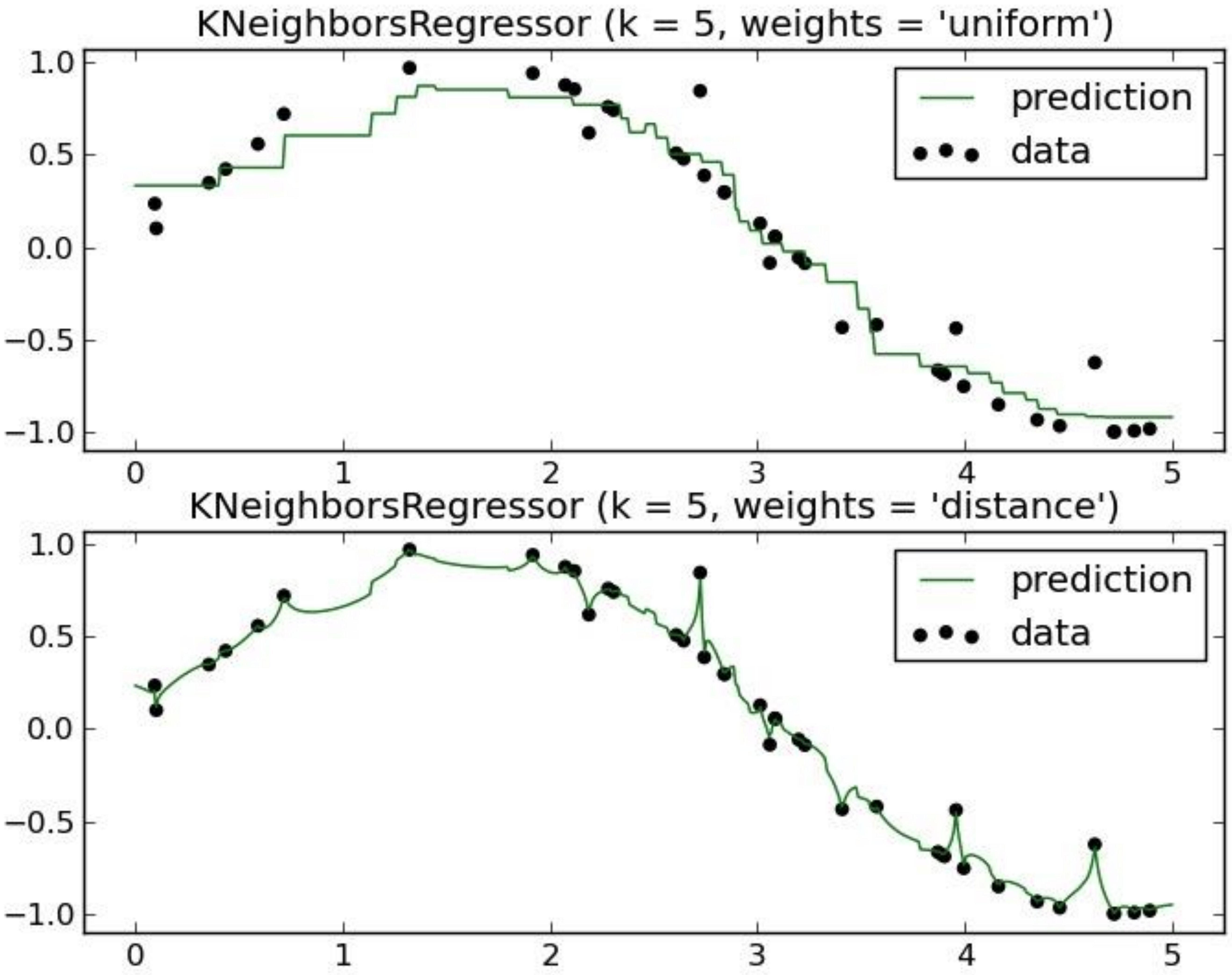
局部回归

- 局部回归（Local regression）尝试利用给定点 x_0 附近的训练样本拟合一个简单的回归模型
  - 得到距离 x_0 最近的 s 个训练样本点，记为 $\mathcal{N}(x_0)$
  - 对每个训练样本点 $x_i \in \mathcal{N}(x_0)$, 根据其距 x_0 的距离赋予相应的权重 $K_h(x_i, x_0)$, 这里 h 表示所考虑近邻的范围
  - 对于那些离 x_0 较远的样本点，其所对应的权重值越小
- 一个例子：K 近邻法

K近邻法

- 考虑样本 $Z^n = \{(x_i\,,\ y_i)\}_{i=1}^n$，其中 $x_i \in R^p\,, y_i \in R.$
- 给定 k 以及一个新样本点 x_0，k近邻法进行如下预测
$$\hat{f}(x_0) = Average(y_i | x_i \in \mathcal{N}_k(x_0)), \qquad (3)$$
- 其中 $\mathcal{N}_k(x_0)$ 表示在欧式距离下距离 x_0 最近的 k 个点
- 这里(3)被用来当作对 $E[y|x]$ 的估计
- 在上述方法中，所有 $\mathcal{N}_k(x_0)$中的点都有着相同的权重
- 故根据 knn 所得到的估计在局部是一个常数，并且整体来看是非连续且复杂度较高
- 想一想：我们将 knn 称为惰性学习方法的原因是什么?
- 如果给 $\mathcal{N}_k(x_0)$ 中的点赋予不同的权重会有怎么样的效果?

K近邻法


KNeighborsRegressor (k = 5, weights = 'uniform')
prediction
data
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5
KNeighborsRegressor (k = 5, weights = 'distance')
prediction
data
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5

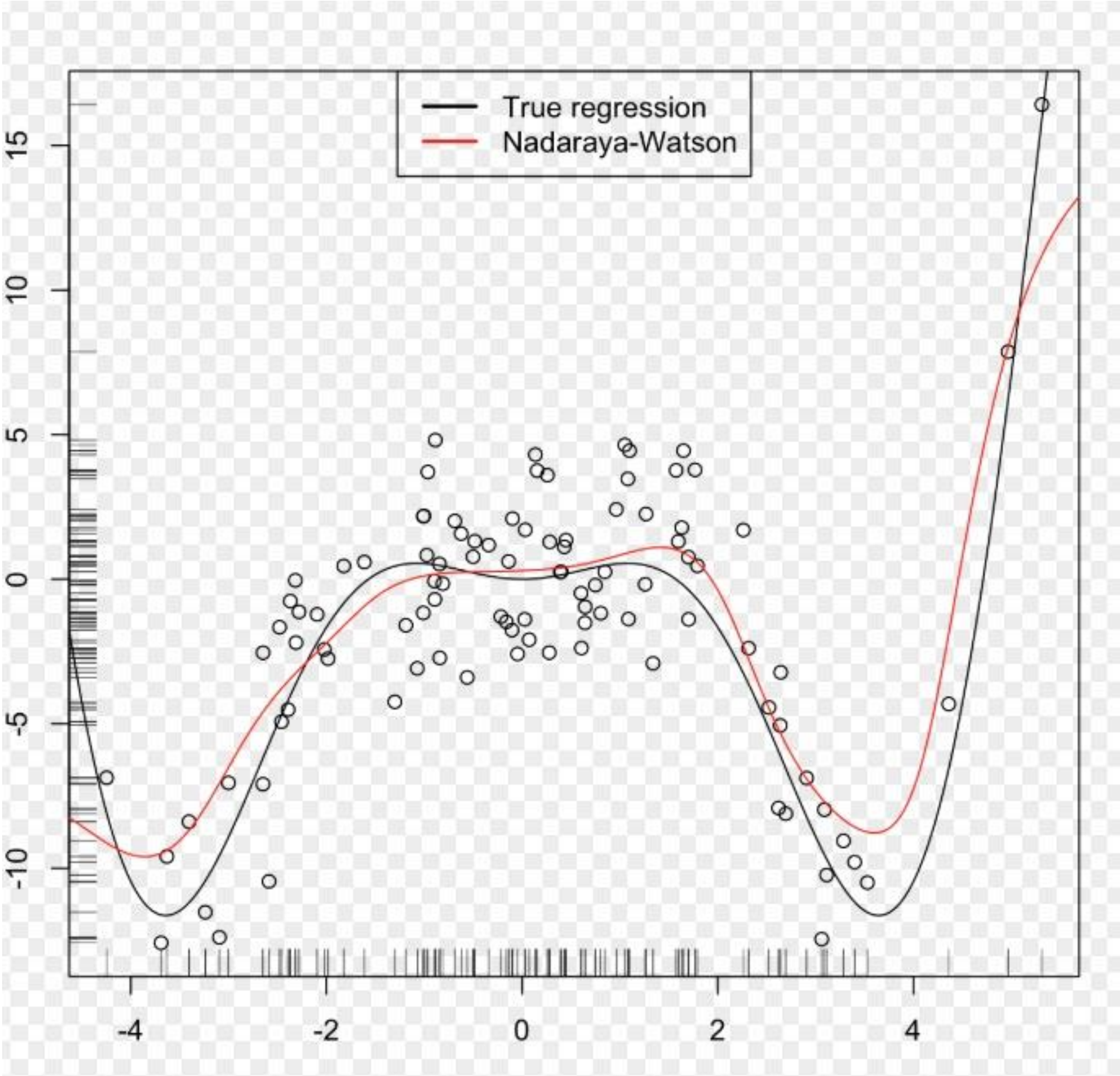
NADARAYA-WATSON 方法

- Nadaraya-Watson (NW) 方法提供了一种对邻域内的点赋予不同权重，进而达到平滑拟合的有效方法
- 给定样本 $Z^n = \{(x_i, y_i)\}_{i=1}^n$ 以及新样本点 x_0
- 考虑一维情况即 $x \in R$，有关NW 方法的预测值为 $\hat{f}(x_0) = \dfrac{\sum_{i=1}^n K_h(x_0, x_i) y_i}{\sum_{i=1}^n K_h(x_0, x_i)}$
- 上式中，一个可选的权重函数为 $K_h(x_0, x) = D(\frac{|x - x_0|}{h(x_0)})$，

$$D(t) = \begin{cases} \frac{3}{4}(1 - t^3) & \text{如果}|t| \leq 1 \\ 0 & \text{如果}|t| > 1 \end{cases}$$

- 上述核函数被称为 Epanechnikov quadratic 核函数
- 在边界附近时，由于样本量较少，因而拟合会显得十分陡峭

NADARAYA-WATSON 方法



NADARAYA-WATSON 方法

- 一些其他的核函数，如
  - Tri-cube 核函数：

$$D(t)=\begin{cases}(1-|t|^3)^3 & \text{如果}|t|\leq 1\\ 0 & \text{如果}|t|>1\end{cases}$$

  - 高斯核密度核函数：

$$K_h(x_0,x)=\frac{1}{h}\exp\Big\{-\frac{|x-x_0|^2}{2h}\Big\}.$$

NADARAYA-WATSON 方法

- 考虑一维情况 $x \in R$，有关NW 方法在新样本点 x_0 的预测值为

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} K_h(x_0, x_i) y_i}{\sum_{i=1}^{n} K_h(x_0, x_i)}$$

- 上式等价于求解如下问题

$$\hat{f} = \operatorname{argmin} \sum_{i=1}^{n} K_h(x_0, x_i)(y_i - f_\theta(x_i))^2,$$

其中$f_\theta(x_i) = \beta_0$.

- 更一般化地，如果考虑线性拟合

$$f_\theta(x_i) = \beta_0 + \beta x_i$$

即为局部线性回归

局部线性回归

- 给定 x_0，考虑如下带有权重项的优化问题

$$\min_{\beta_0,\beta_1}\sum_{i=1}^{n} K_h(x_0,x_i)(y_i-\beta_0-\beta_1 x_i)^2$$

- 记 $b(x)=(1,x)^T, \beta=(\beta_0,\beta_1)^T$.
- $B=\left(b(x_1),\dots,b(x_n)\right)^T\in R^{n\times 2}, W(x_0)=Diag\left(K_h(x_0,x_1),\dots,K_h(x_0,x_n)\right)^T\in R^{n\times n}$
- 上述问题可以写为向量形式

$$\min_{\boldsymbol{\beta}}(Y-B\boldsymbol{\beta})^T W(x_0)(Y-B\boldsymbol{\beta})$$

- 直接计算可得

$$\widehat{\boldsymbol{\beta}}=\left(B^T W(x_0) B\right)^{-1} B^T W(x_0) Y$$

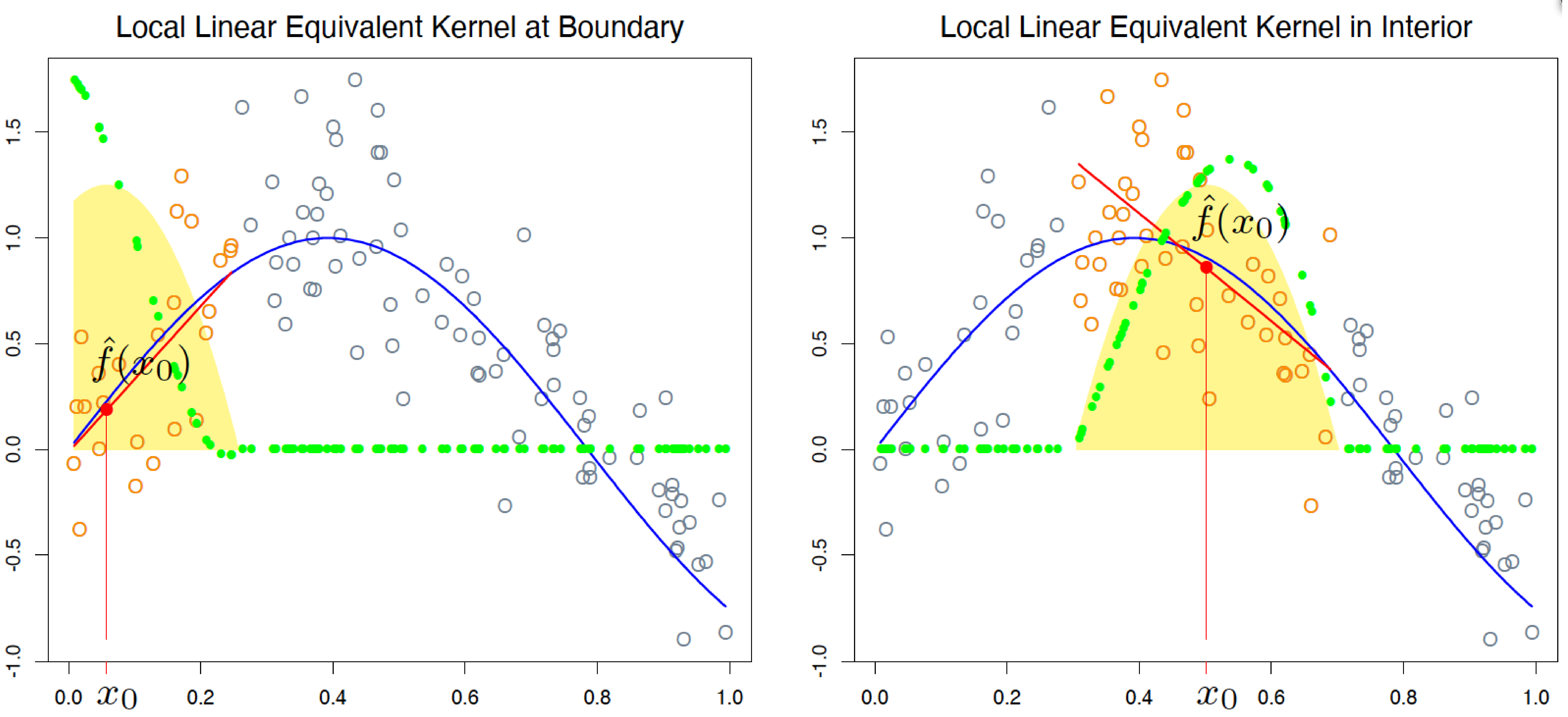
局部线性回归

- 在 x_0 点处的拟合值为

$$\begin{aligned}\hat{f}(x_0) &= b(x_0)^T\widehat{\boldsymbol{\beta}} \\ &= b(x_0)^T\left(B^TW(x_0)B\right)^{-1}B^TW(x_0)Y \\ &= \sum_{i=1}^{n}\ell_i(x_0)y_i,\end{aligned}$$

- 可见 $\hat{f}(x_0)$是有关 y_i 的线性函数
- $l_i(x_0)$ 也被称为equivalent kernel.

局部线性回归


Local Linear Equivalent Kernel at Boundary
Local Linear Equivalent Kernel in Interior
$\hat{f}(x_0)$
x_0
1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0

局部多项式回归

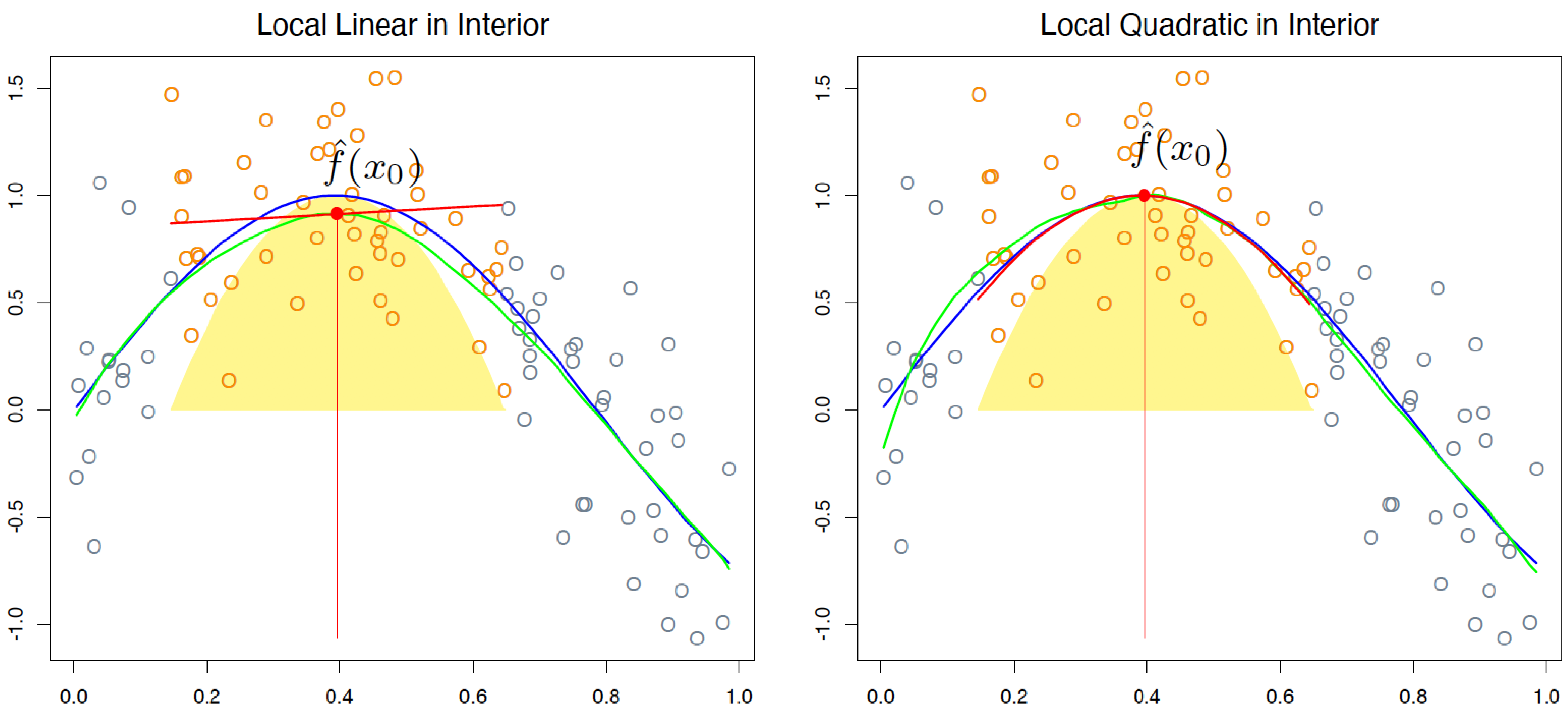
- 给定 x_0，考虑如下带有权重项的最小二乘

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} K_h(x_0, x_i)(y_i - \beta_0 - \sum_{j=1}^{d} \boldsymbol{\beta}_j^T x_i^j)^2$$

- 因而，在x_0点处的拟合值为

$$\hat{f}(x_0) = \hat{\beta}_0 + \sum_{j=1} \hat{\boldsymbol{\beta}}_j^T x_0^j$$

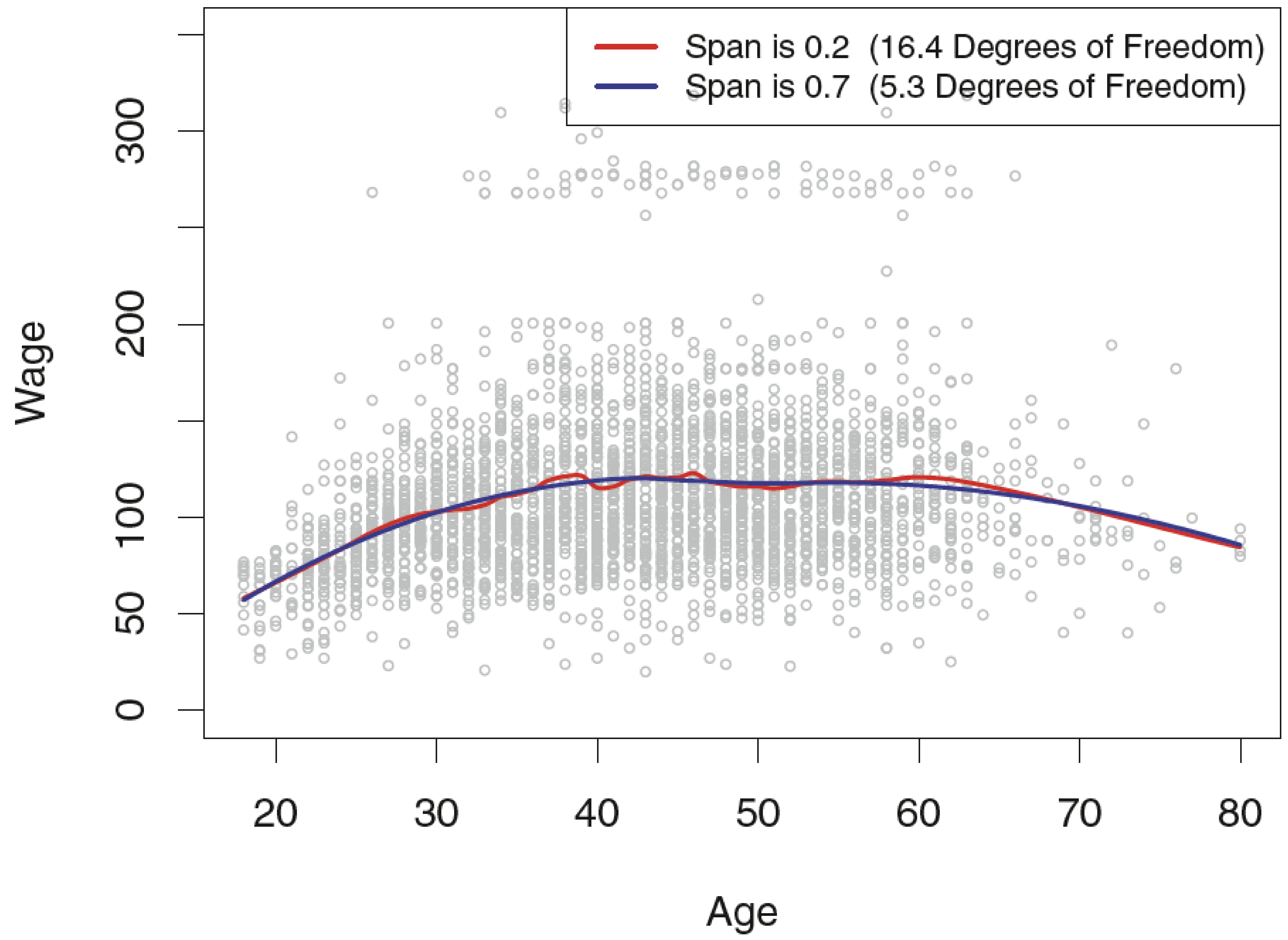
局部多项式回归


Local Linear in Interior
Local Quadratic in Interior
$\hat{f}(x_0)$
$\hat{f}(x_0)$
1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0

注意事项

- 有关权重函数 K 的选取
- 有关所拟合函数的选取:局部常数，线性或二次/多项式局部回归函数
- 有关邻域范围（span）的选取
  - 它控制了非线性拟合的灵活性
  - 较小的邻域会导致更加的局部以及较大摆动的拟合，而较大的邻域通过使用所有训练观测值会导致全局拟合
  - 可以通过交叉验证进行选取

一个局部回归的例子


Span is 0.2 (16.4 Degrees of Freedom)
Span is 0.7 (5.3 Degrees of Freedom)
Wage
0
50
100
200
300
Age
20
30
40
50
60
70
80

思考

- 上面我们所考虑的都是基于一维情况下的非线性拟合，上述方法均可扩展到多变量的情况下
- 可加模型假设提供了一种针对多变量问题建立非参数统计学模型的一般化框架

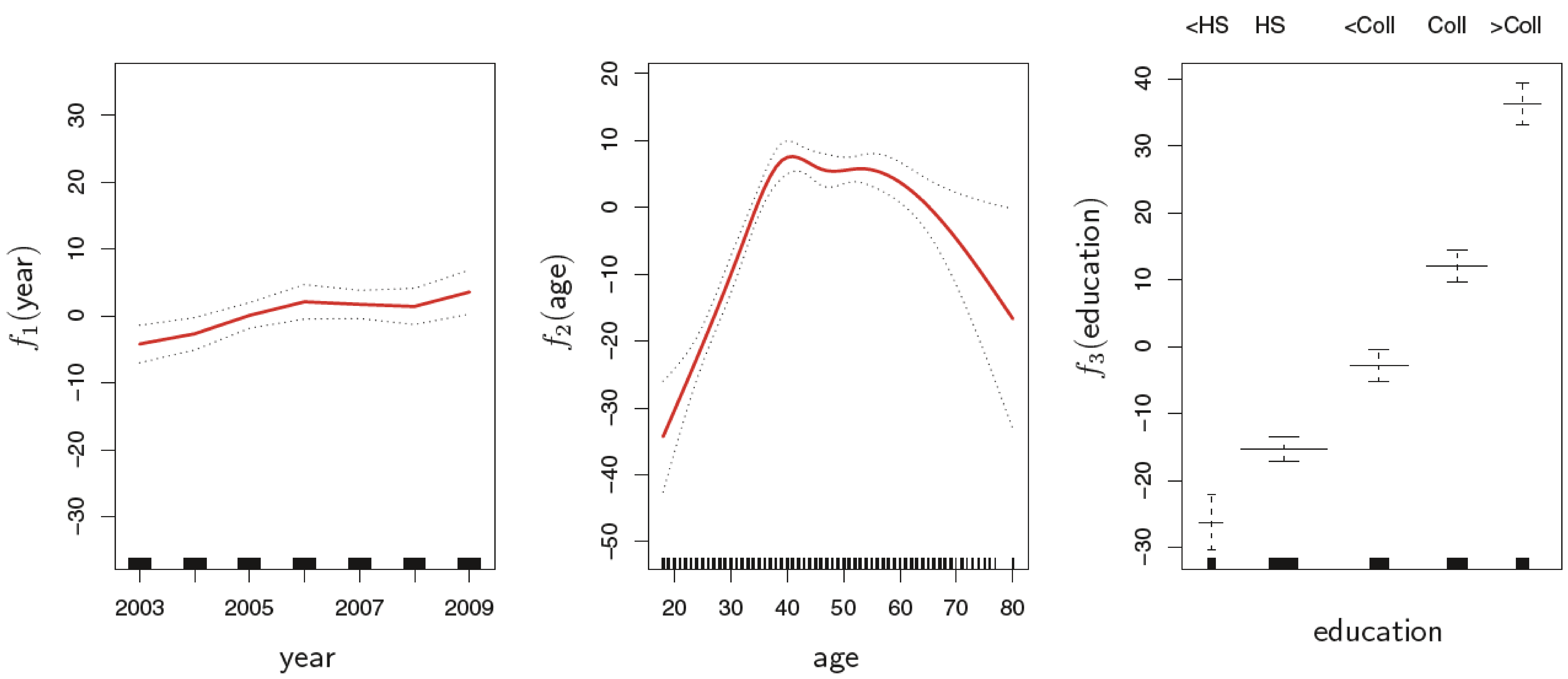
可加模型

- 可加模型（Additive model）提供了一个可以针对多变量问题建立非参数统计学模型的一般化框架
- 该模型假设

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

  其中每一个 f_j 都是变量 x_j 的一个函数
- f_j 可以是任何形式的函数：线性或非线性.

可加模型


$f_1(\text{year})$
30
20
10
0
−10
−20
−30
2003
2005
2007
2009
year
$f_2(\text{age})$
20
10
0
−10
−20
−30
−40
−50
20
30
40
50
60
70
80
age
<HS
HS
<Coll
Coll
>Coll
$f_3(\text{education})$
40
30
20
10
0
−10
−20
−30
education

优势与缺点

- 可加模型对每一个 x_j 拟合一维的非线性函数 f_j，并将这 p 个一维非线性函数相加以建立统计模型
- 该模型本质上是可加的，因此我们在保持所有其他变量不变的情况下检查每个单独的 x_j 对 Y 的影响
- 可加模型依然是一个有着较强限制性条件的模型假设，其忽略了变量之间可能存在的交互作用
- 如果在建模过程中需要考虑变量之间的交互作用的影响，可以考虑将交叉项 $f_{jk}(x_j, x_k)$ 引入可加模型中

思考

- 如何利用所学到的上述知识具体拟合可加模型呢?