Nonparametric Regression

Nihar Shah

DSSG

9 July 2013

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$
2. Polynomial regression $y = \beta_0 + \beta_1 x + \beta_2 x^2$

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$
2. Polynomial regression $y = \beta_0 + \beta_1 x + \beta_2 x^2$
3. Logit regression $y = \frac{1}{1+e^{\beta_0+\beta_1 x}}$

What do these have in common?

1. Linear regression $$y = \beta_0 + \beta_1 x$$
2. Polynomial regression $$y = \beta_0 + \beta_1 x + \beta_2 x^2$$
3. Logit regression $$y = \frac{1}{1+e^{\beta_0+\beta_1 x}}$$
4. Probit regression $$y = \Phi(\beta_0 + \beta_1 x)$$

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$
2. Polynomial regression $y = \beta_0 + \beta_1 x + \beta_2 x^2$
3. Logit regression $y = \frac{1}{1+e^{\beta_0+\beta_1 x}}$
4. Probit regression $y = \Phi(\beta_0 + \beta_1 x)$
5. Poisson regression $y = e^{\beta_0+\beta_1 x}$

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$
2. Polynomial regression $y = \beta_0 + \beta_1 x + \beta_2 x^2$
3. Logit regression $y = \frac{1}{1+e^{\beta_0+\beta_1 x}}$
4. Probit regression $y = \Phi(\beta_0 + \beta_1 x)$
5. Poisson regression $y = e^{\beta_0+\beta_1 x}$
6. Quantile regression $y_\tau = \beta_0 + \beta_1 x$

What do these have in common?

1. Linear regression $y = \beta_0 + \beta_1 x$
2. Polynomial regression $y = \beta_0 + \beta_1 x + \beta_2 x^2$
3. Logit regression $y = \frac{1}{1+e^{\beta_0+\beta_1 x}}$
4. Probit regression $y = \Phi(\beta_0 + \beta_1 x)$
5. Poisson regression $y = e^{\beta_0+\beta_1 x}$
6. Quantile regression $y_\tau = \beta_0 + \beta_1 x$

These all assume a model!

By contrast...

Nonparametric regression makes no model assumptions (i.e. no parameters to estimate)

$$y = f(x; x_1, x_2, x_3, ..., x_n)$$

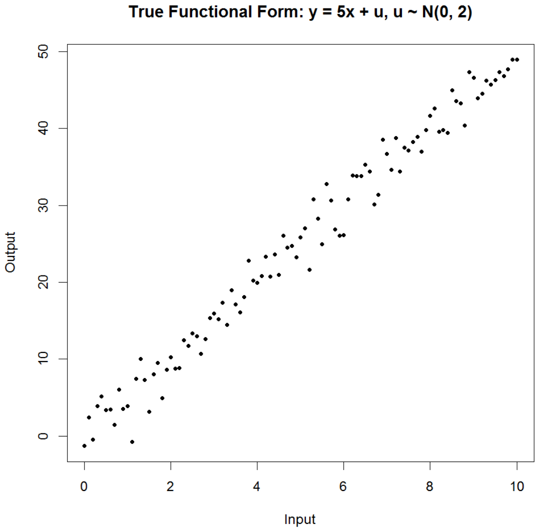
By contrast...

Nonparametric regression makes no model assumptions (i.e. no parameters to estimate)

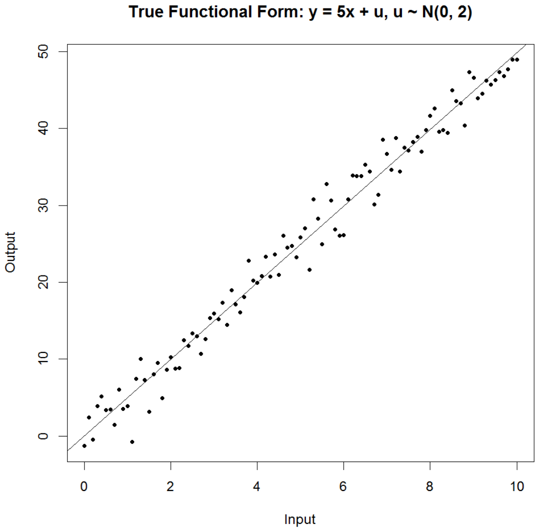
$$y = f(x; x_1, x_2, x_3, ..., x_n)$$

- Parametric estimation: data $\Rightarrow$ parameters $\Rightarrow$ estimation
- Nonparametric estimation: data $\Rightarrow$ estimation

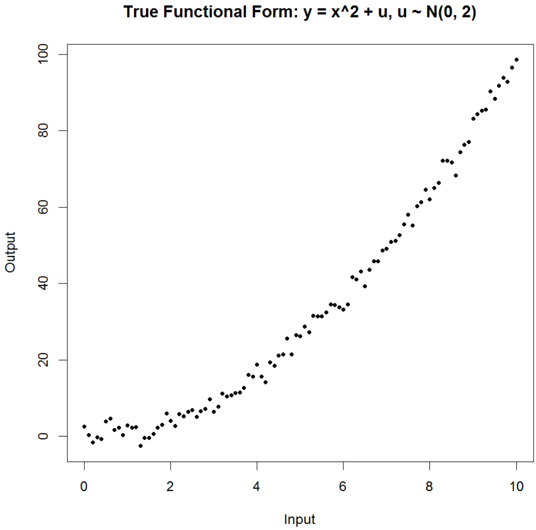
Parametric regression works great sometimes...



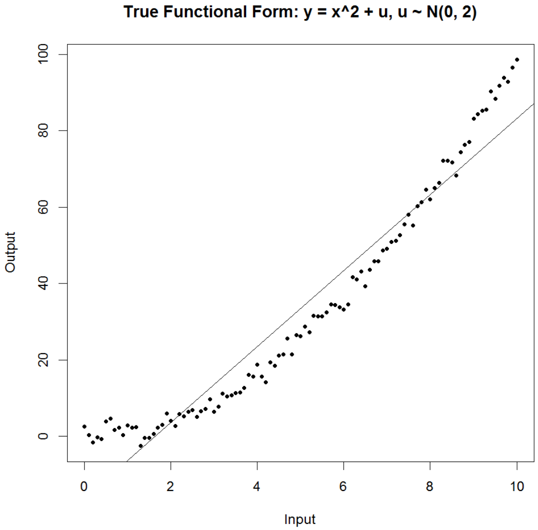
Parametric regression works great sometimes...



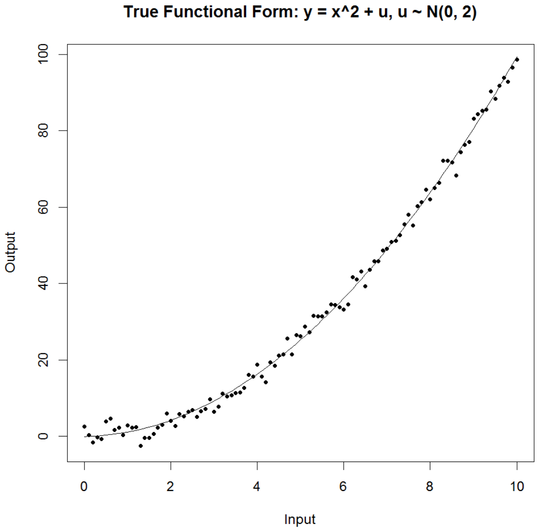
Parametric regression works great sometimes...



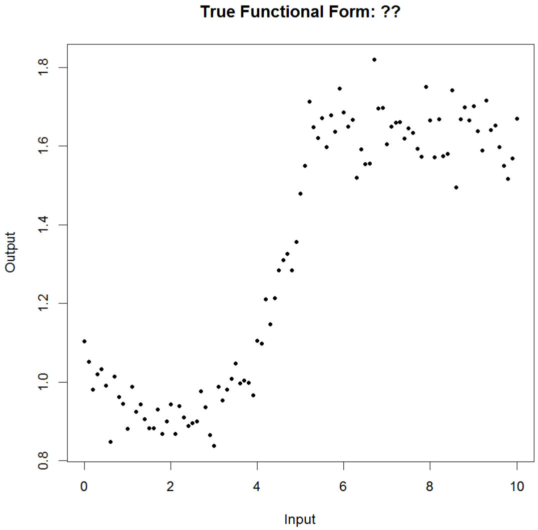
Parametric regression works great sometimes...



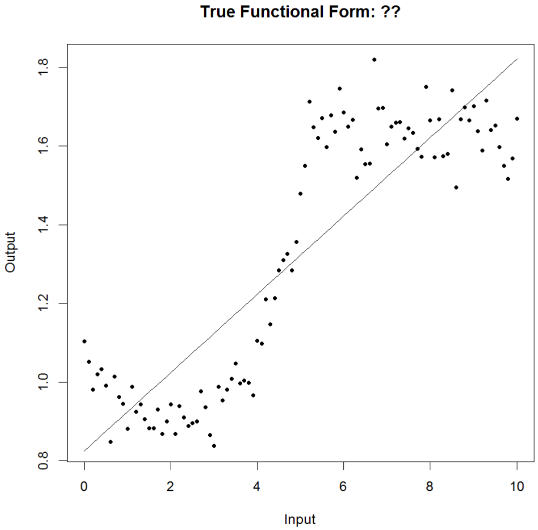
Parametric regression works great sometimes...



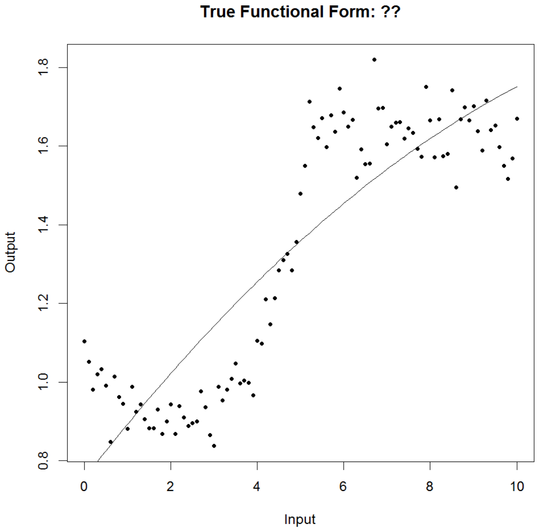
...but not all the time



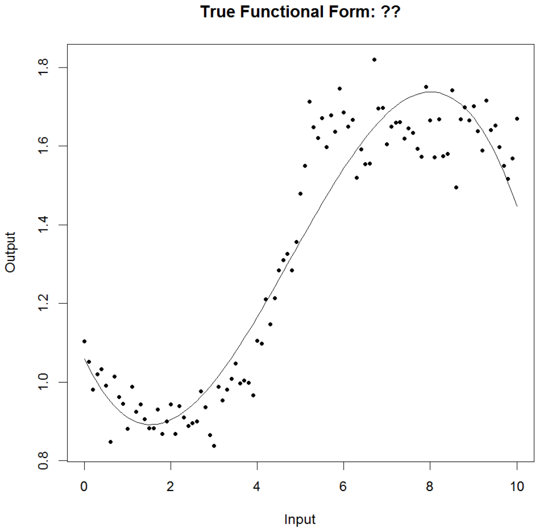
...but not all the time



...but not all the time

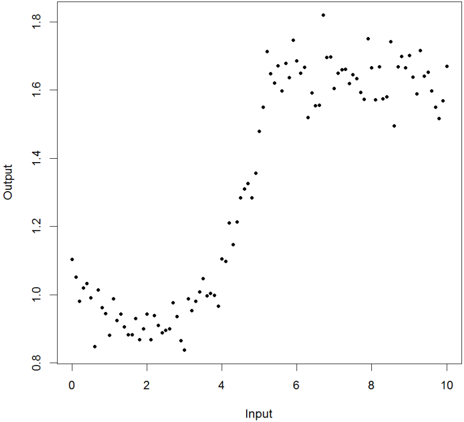


...but not all the time



...but not all the time

$$1.02^{\min\{0.5x^3-1.5x^2+1.1x-6\sqrt{x}+3,\ 25\}+u},\ u \sim \mathcal{N}(0,2)$$



Types of Nonparametric Regression

1. Kernel regression
2. Nonparametric multiplicative regression
3. Regression trees

Types of Nonparametric Regression

1. $\boxed{\text{Kernel regression}}$
2. Nonparametric multiplicative regression
3. Regression trees

Overview

Kernel nonparametric regression is just...

Overview

Kernel nonparametric regression is just... weighted averages!

Overview

Kernel nonparametric regression is just... weighted averages!

Say you want to predict a response ($\hat{y}$) at a given point (x)

- Linear regression
  - Use available data $(x_1, y_1), (x_2, y_2), (x_3, y_3)$... to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$
  - Use $\hat{\beta}_0$ and $\hat{\beta}_1$ to now predict $\hat{y}$, or $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Overview

Kernel nonparametric regression is just... weighted averages!

Say you want to predict a response ($\hat{y}$) at a given point (x)

- Linear regression
  - Use available data $(x_1, y_1), (x_2, y_2), (x_3, y_3)$... to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$
  - Use $\hat{\beta}_0$ and $\hat{\beta}_1$ to now predict $\hat{y}$, or $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- Kernel regression
  - Calculate a weight for each data point: w_1 for (x_1, y_1), w_2 for (x_2, y_2), w_3 for (x_3, y_3), etc.
  - Estimate $\hat{y}$ as a weighted average of the responses: $\sum w_i y_i$

Overview

Kernel nonparametric regression is just... weighted averages!

Say you want to predict a response ($\hat{y}$) at a given point (x)

- Linear regression
  - Use available data $(x_1, y_1), (x_2, y_2), (x_3, y_3)$... to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$
  - Use $\hat{\beta}_0$ and $\hat{\beta}_1$ to now predict $\hat{y}$, or $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- Kernel regression
  - Calculate a weight for each data point: w_1 for (x_1, y_1), w_2 for (x_2, y_2), w_3 for (x_3, y_3), etc.
  - Estimate $\hat{y}$ as a weighted average of the responses: $\sum w_i y_i$
  - The weight is computed by the Kernel

Heights versus Weight

Height (in)	Weight (lb)
68	140
70	150
72	165
74	170

Now say we want to estimate the weight of someone who is 71 inches.

Heights versus Weight

Height (in)	Weight (lb)
68	140
70	150
72	165
74	170

Now say we want to estimate the weight of someone who is 71 inches.

Linear regression: $\hat{\beta}_0 = -216.5, \hat{\beta}_1 = 5.25$
Estimation: $\hat{y} = -216.5 + 5.25 \cdot 71 = 156.25$

Heights versus Weight

Take a made-up kernel: $K(x_1, x_2) = \frac{1}{|x_1 - x_2|}$

$x = 71$ inches, $\hat{y} =?$

Height (in)	Weight (lb)	Original Weight	Normalized Weight
68	140	0.333	0.125
70	150	1	0.375
72	165	1	0.375
74	170	0.333	0.125

$\hat{y} = 0.125 \cdot 140 + 0.375 \cdot 150 + 0.375 \cdot 165 + 0.125 \cdot 170$

$\hat{y} = 156.875$

Theory

Nadaraya-Watson Kernel Estimator

$$\boxed{\hat{y} = \widehat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i) Y_i}{\sum_{i=1}^{n} K_h(x - X_i)}}$$

Popular Kernels

- Uniform
- Triangular
- Normal
- Epanechnikov

Kernels: Some Details

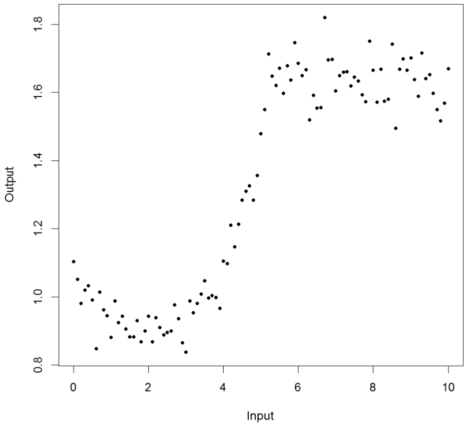
Turns out they are a little more complicated than just weights...

- They have a parameter (h) called bandwidth
- They can be multiplied when we have several predictors (e.g. height and gender):

$$K_h(x - X_i) = K_h^1(x^1 - X_i^1)K_h^2(x^2 - X_i^2)$$

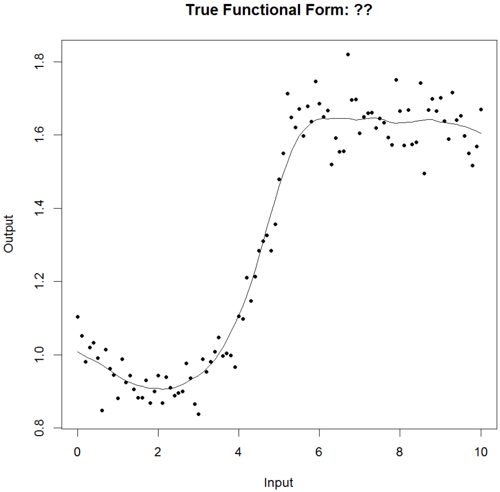
Returning to Original Plot

$$1.02^{\min\{0.5x^3 - 1.5x^2 + 1.1x - 6\sqrt{x} + 3,\ 25\} + u}, u \sim \mathcal{N}(0, 2)$$



Returning to Original Plot

Using an Epanechnikov Kernel...

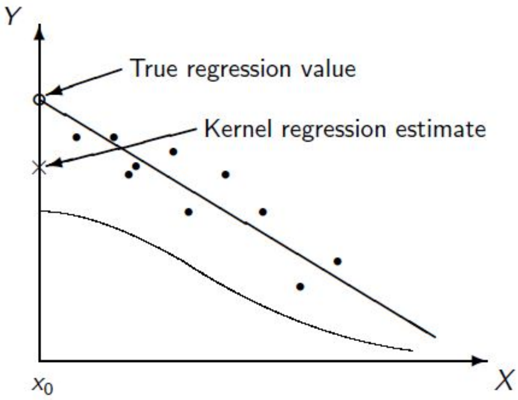


Why not Nonparametric Regression?

Why not Nonparametric Regression?

- Extrapolation is problematic

Why not Nonparametric Regression?



Why not Nonparametric Regression?

- Extrapolation is problematic

Why not Nonparametric Regression?

- Extrapolation is problematic
- Prediction is computationally intensive

Why not Nonparametric Regression?

- Extrapolation is problematic
- Prediction is computationally intensive
- No parameters to analyze

Why not Nonparametric Regression?

- Extrapolation is problematic
- Prediction is computationally intensive
- No parameters to analyze
- Requires lots of data for accurate prediction

Why not Nonparametric Regression?

- Extrapolation is problematic
- Prediction is computationally intensive
- No parameters to analyze
- Requires lots of data for accurate prediction
- Still requires some assumptions...