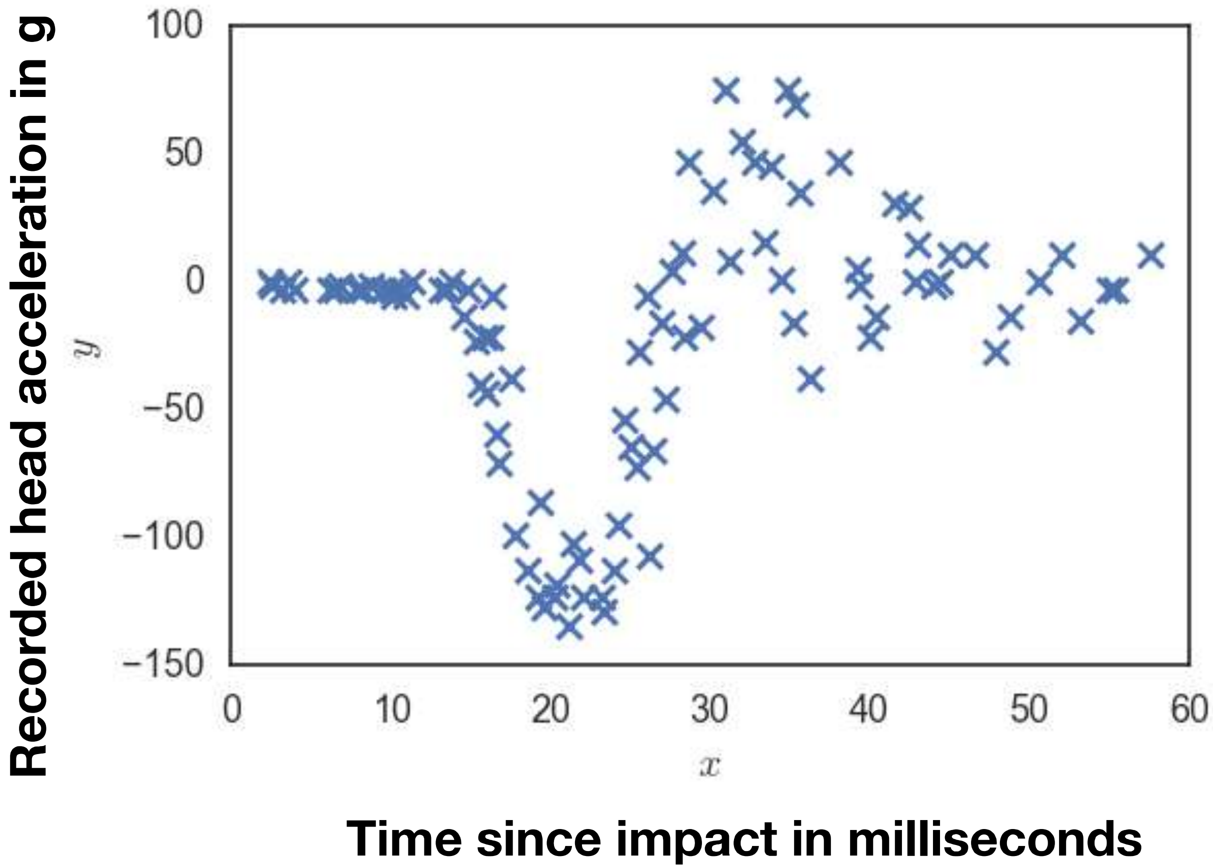
Lecture 13: Linear Regression via Least Squares

Professor Ilias Bilionis

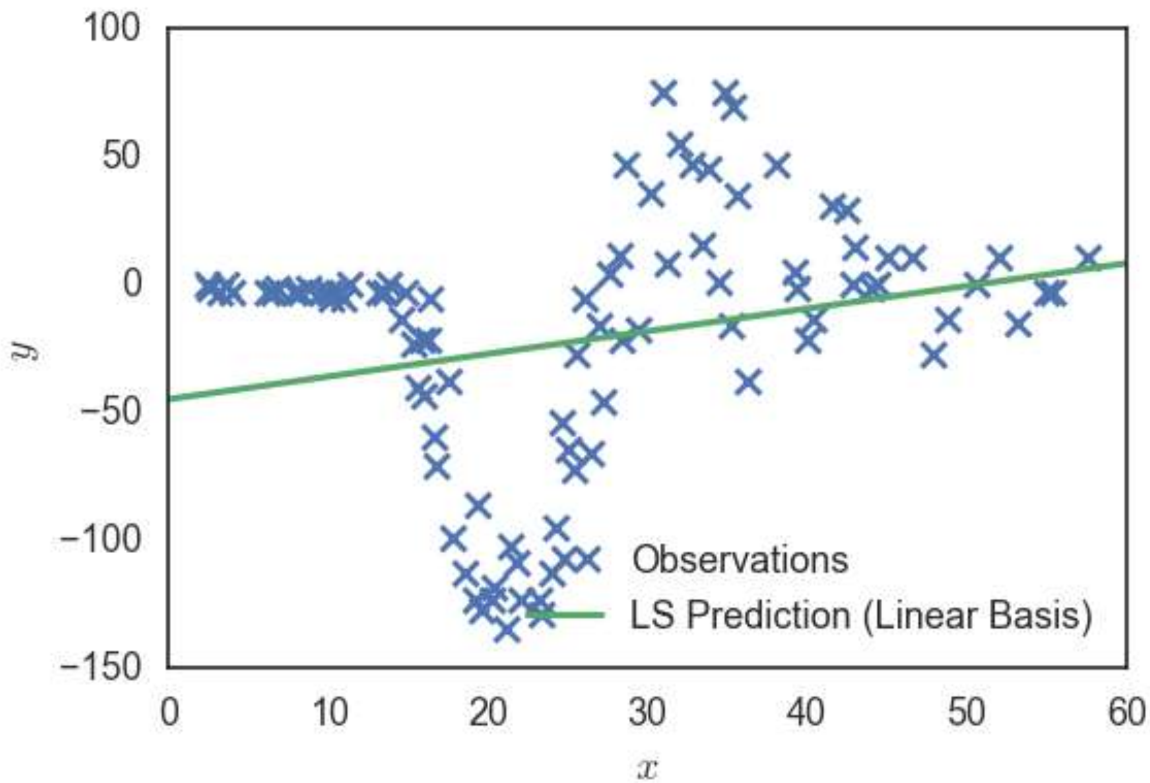
The generalized linear model

PREDICTIVE
SCIENCE LABORATORY

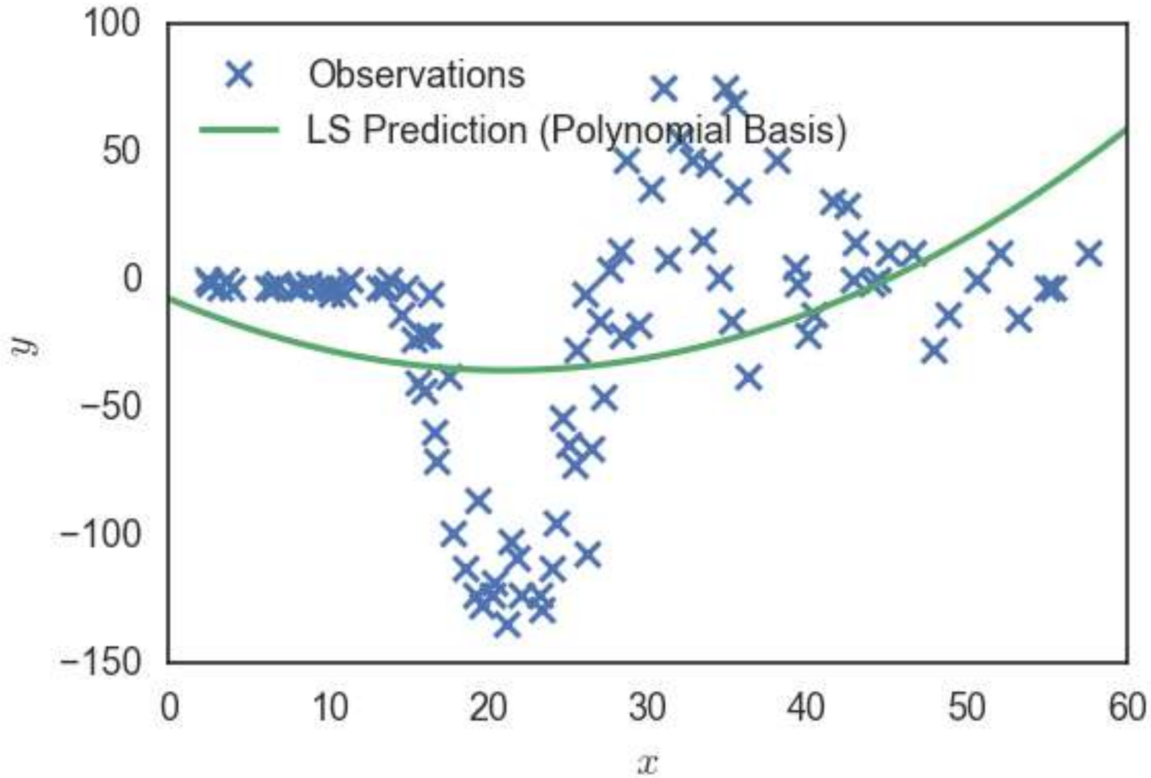
Regression Example (Motorcycle Data Set)



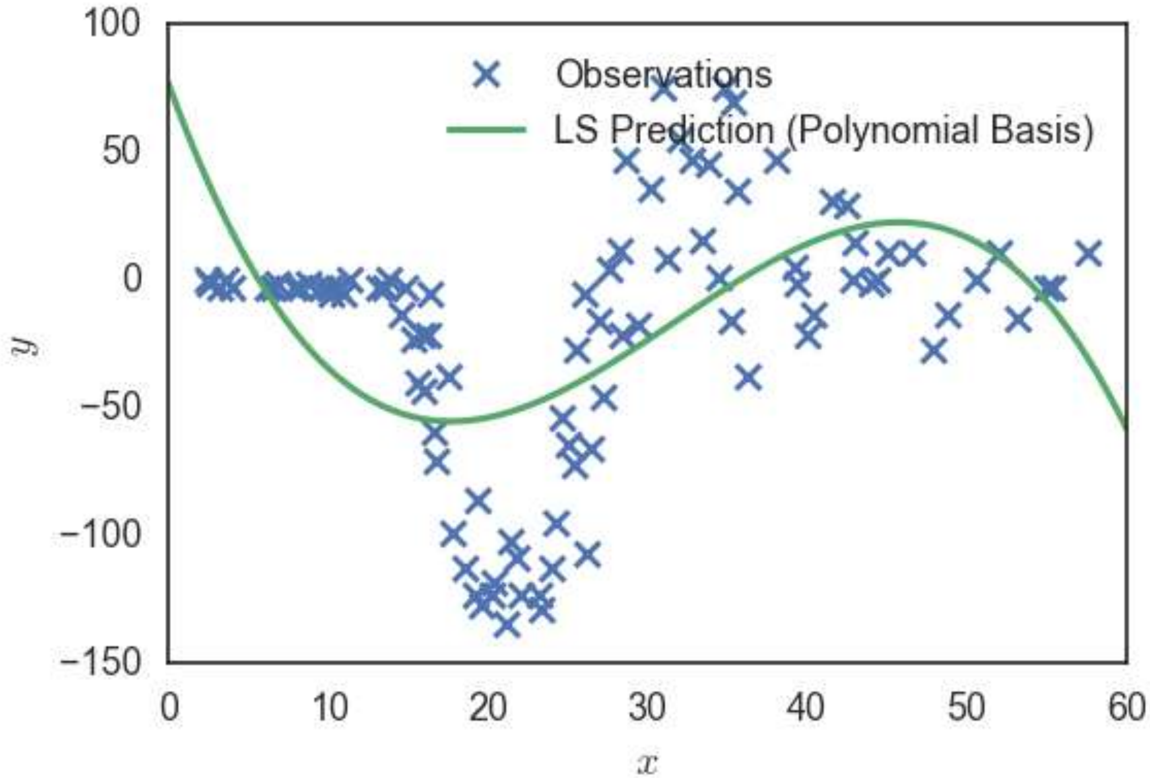
Regression Example: Least Squares with Linear Basis



Regression Example: Least Squares with Polynomial Basis (degree 2)

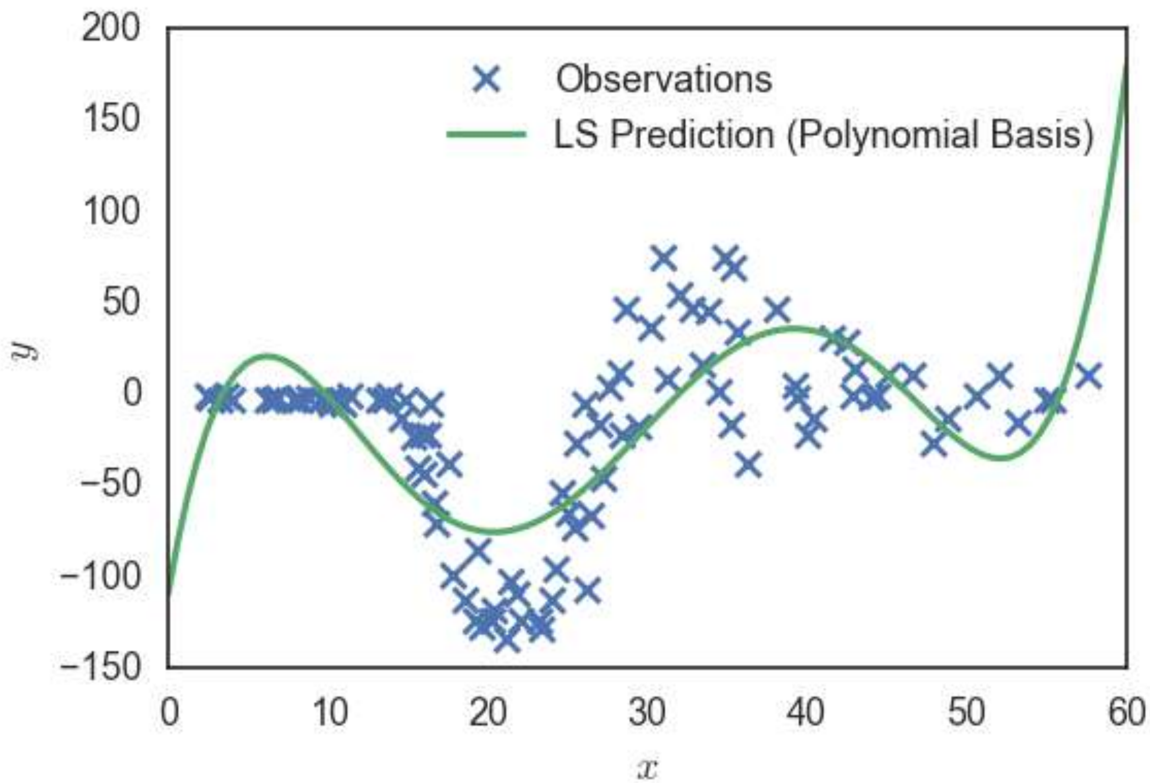


Regression Example: Least Squares with Polynomial Basis (degree 3)



PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Polynomial Basis (degree 5)



PREDICTIVE
SCIENCE LABORATORY

The generalized linear model

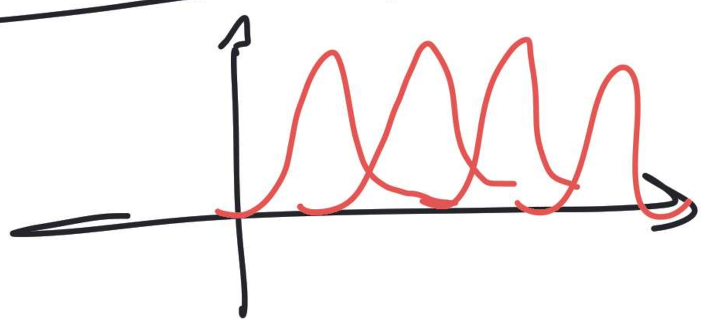
Basis Functions

$$y = w_1 \cdot \varphi_1(x) + w_2 \varphi_2(x) + \dots + w_m \varphi_m(x)$$

Polynomial: $\varphi_1(x) = 1, \quad \varphi_2(x) = x, \quad \varphi_3(x) = x^2, \quad \dots$

Fourier: $\varphi_1(x) = 1, \quad \varphi_2(x) = \cos\left(\frac{2\pi x}{L}\right), \quad \varphi_3(x) = \sin\left(\frac{2\pi x}{L}\right), \dots$

Radial Basis Functions: $\varphi_1(x) = \exp\left\{-\frac{(x - \overbrace{x_{c1}}^{\text{center}})^2}{\ell}\right\}$



Least squares loss function

$$L(w) = \sum_{i=1}^{n}\left[y_i - \left(w_1\varphi_1(x_i) + w_2\varphi_2(x_i) + \dots + w_m\varphi_m(x_i)\right)\right]^2$$

$$= \|\underline{y} - \underline{\underline{\Phi}} \cdot \underline{w}\|_2^2$$

$$\underline{w} = (w_1, \dots, w_m)$$

Design matrix:

$$\underline{\underline{\Phi}} = \begin{pmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_m(x_1) \\ \vdots & \vdots & & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \dots & \varphi_m(x_n) \end{pmatrix}$$

$(n \times m)$

obs

φ's (features)

Minimizing the loss function

$$L(\underline{w}) = \| \underline{y} - \underline{\underline{\Phi}} \cdot \underline{w} \|_2^2$$

$$\nabla_{\underline{w}} L(\underline{w}) = 0$$

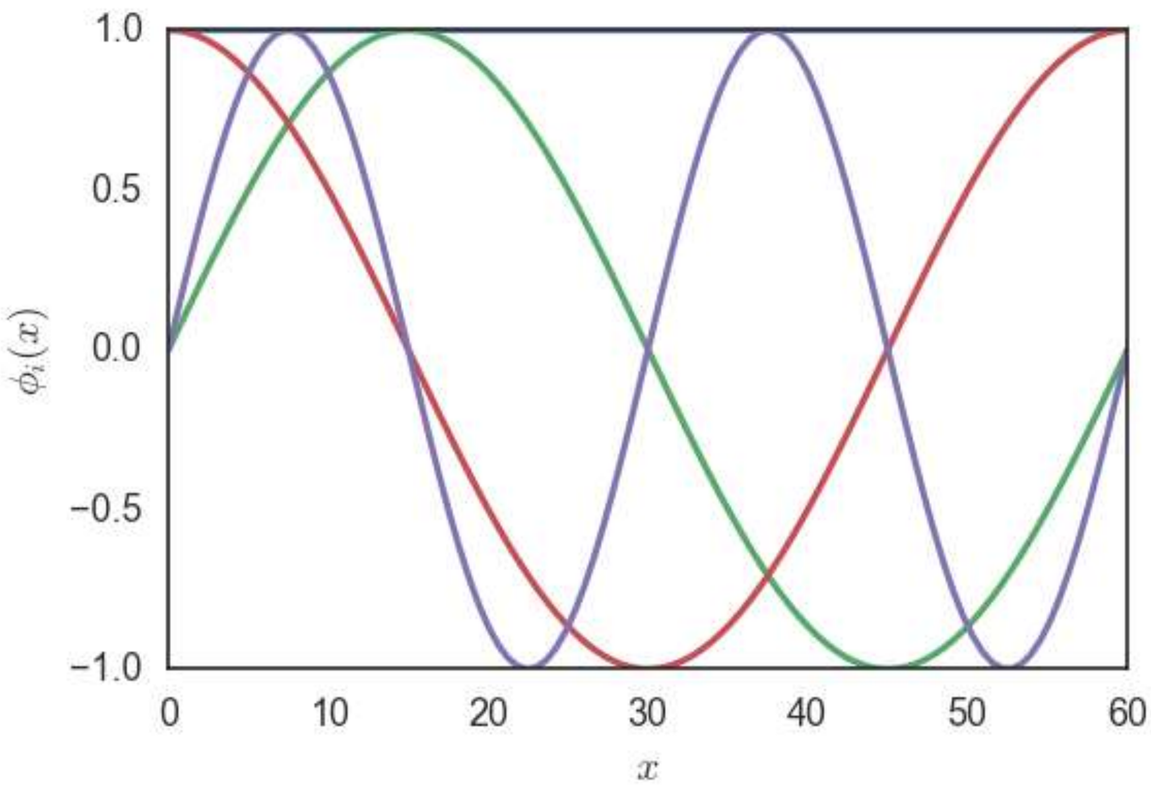
find

$$\Rightarrow \left(\underline{\underline{\Phi}}^T \cdot \underline{\underline{\Phi}} \right) \cdot \underline{w} = \underline{\underline{\Phi}}^T \cdot \underline{y}$$

new notation, same form

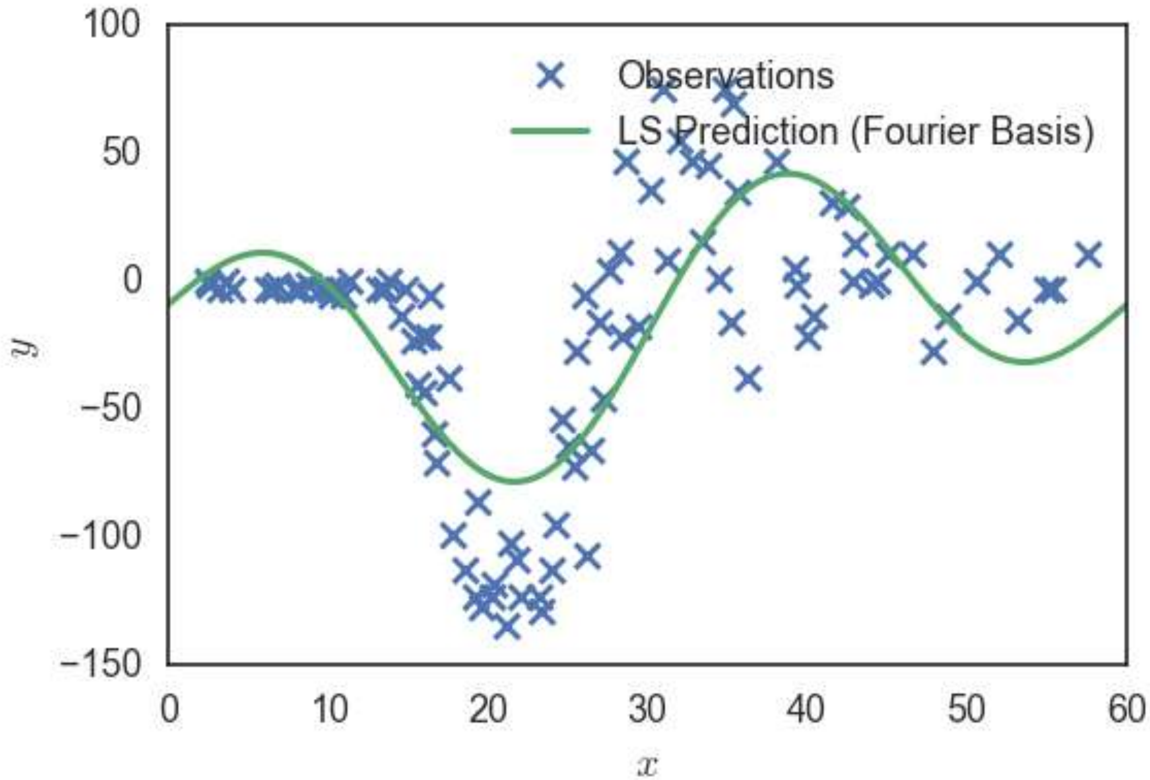
$$\Rightarrow \quad \underline{w} = ?$$

Regression Example: Least Squares with Fourier Basis (4 terms)

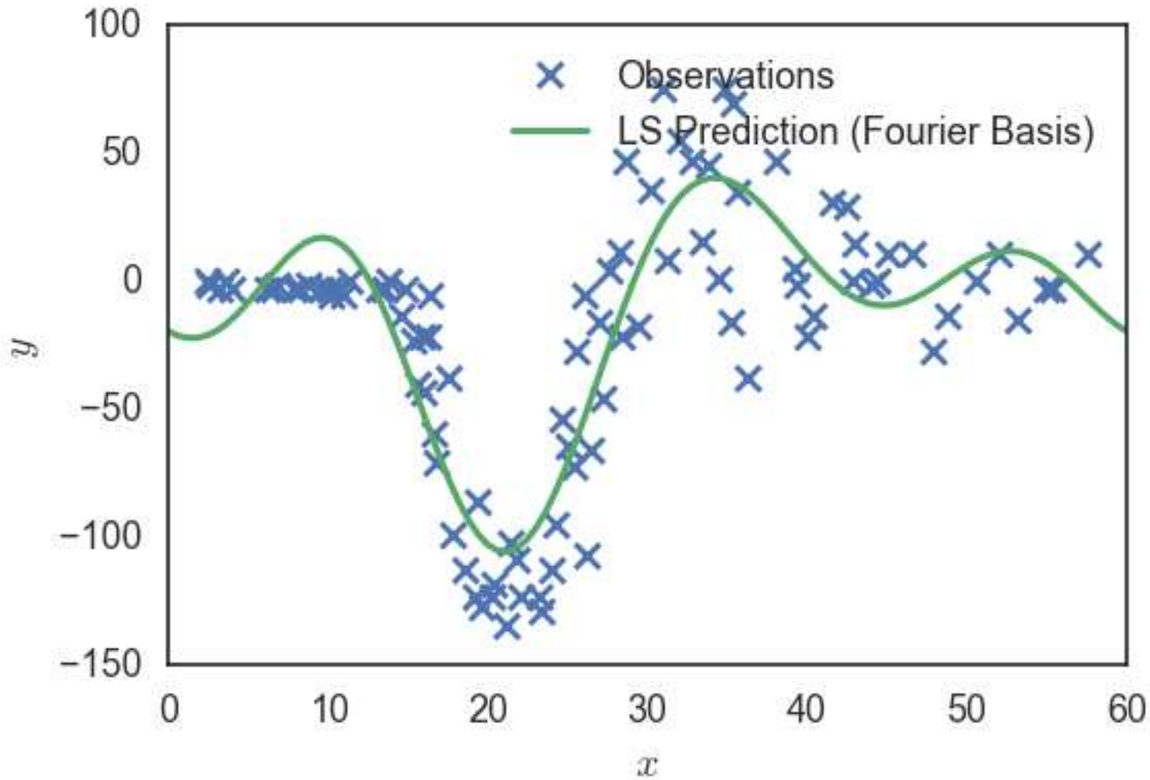


PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Fourier Basis (4 terms)

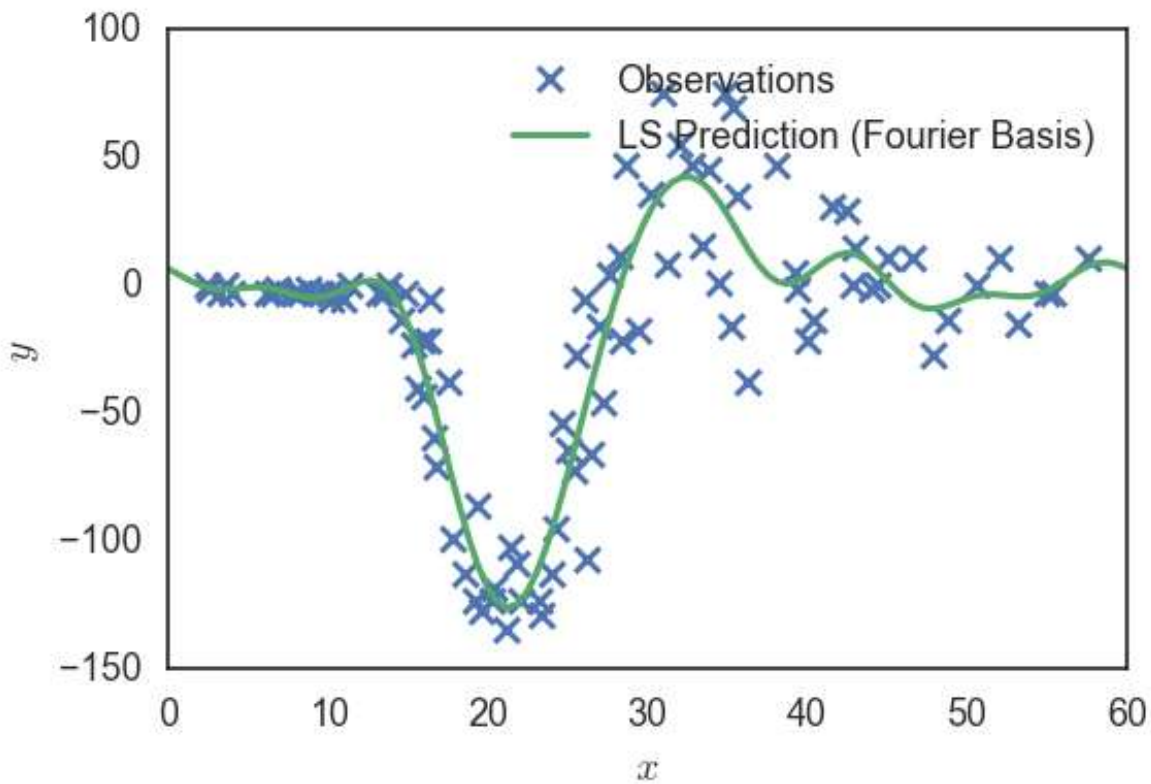


Regression Example: Least Squares with Fourier Basis (8 terms)



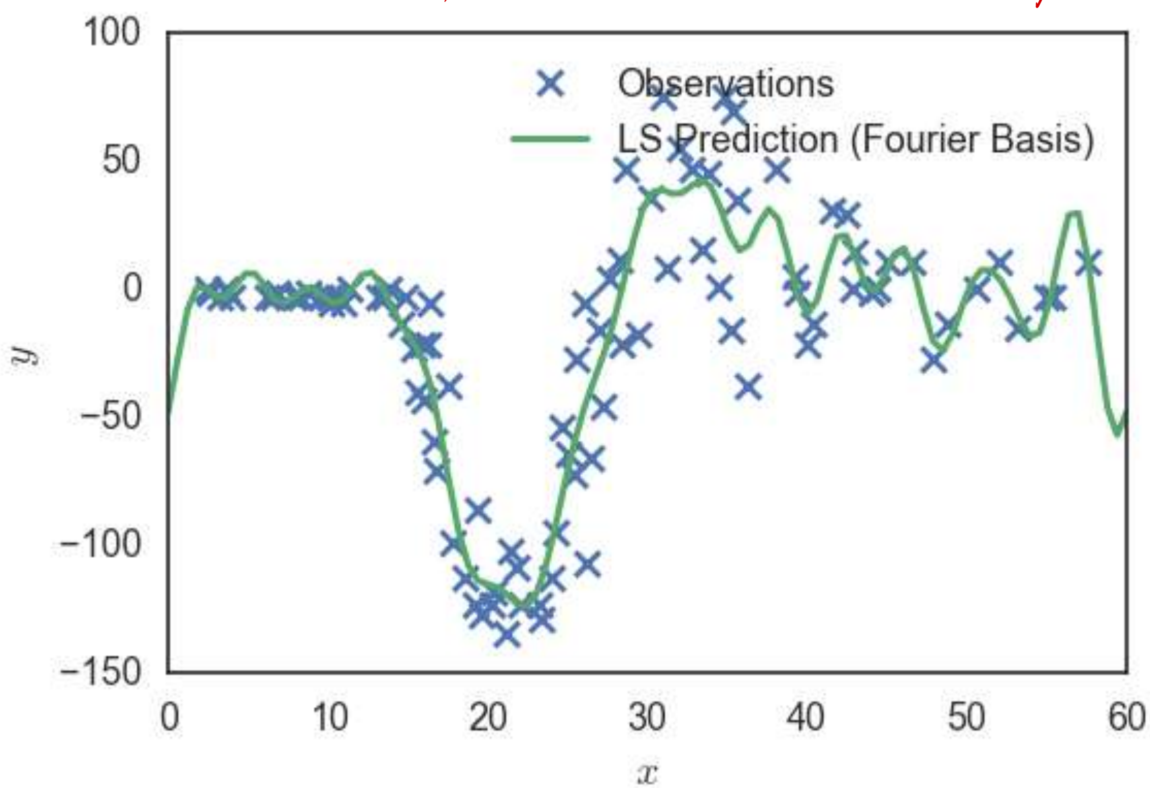
PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Fourier Basis (16 terms)



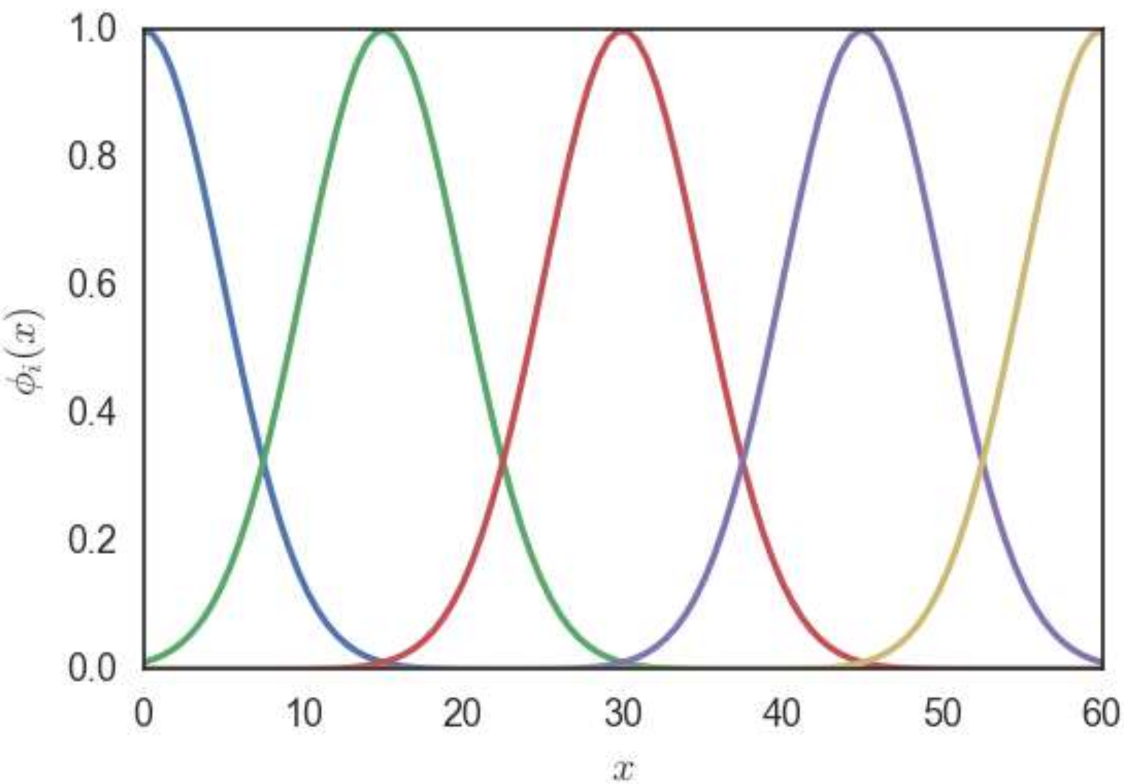
Regression Example: Least Squares with Fourier Basis (32 terms)

Overfit, lots of expressivity



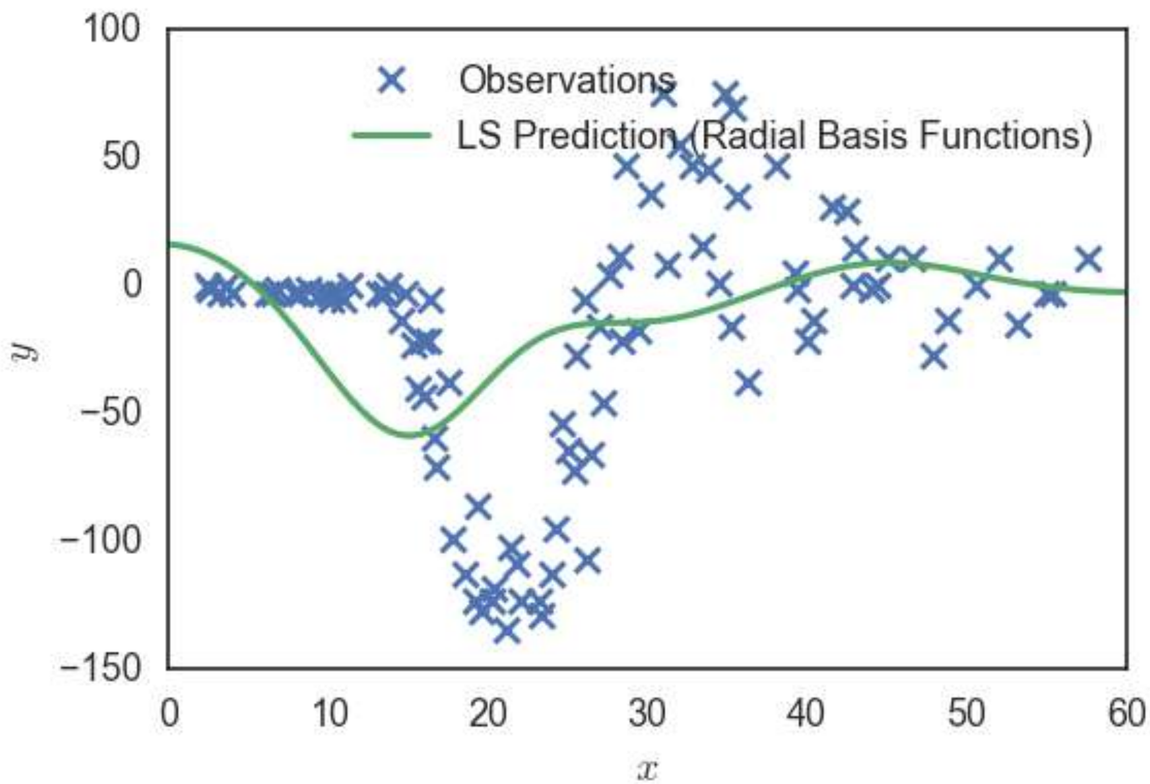
PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Radial Basis (5 terms)



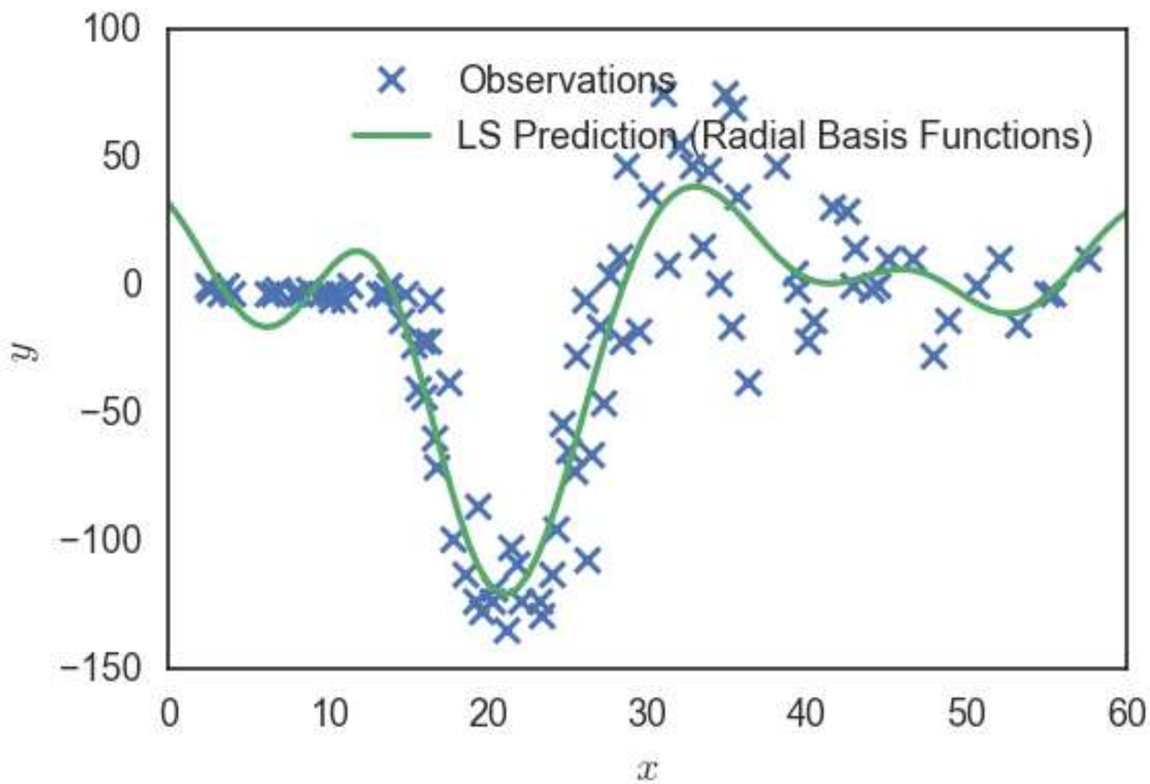
PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Radial Basis (5 terms)



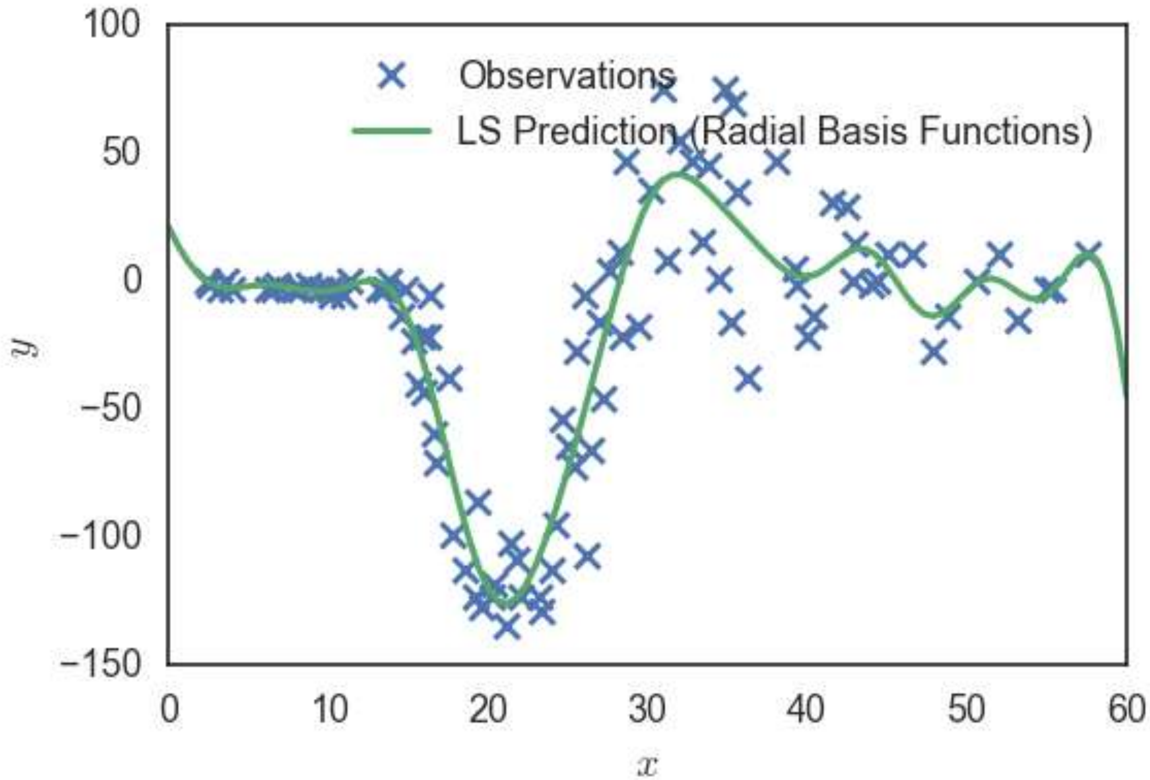
PREDICTIVE
SCIENCE LABORATORY

Regression Example: Least Squares with Radial Basis (10 terms)



Regression Example: Least Squares with Radial Basis (20 terms)

Overfitting is not a basis functions problem, it is a least squares problem



PREDICTIVE SCIENCE LABORATORY

Open questions

- How do I quantify the measurement noise?
- How many basis functions should I use?
- Which basis functions should I use?
- How do I quantify epistemic uncertainty induced by limited data?
- How do I pick the parameters of the basis functions, e.g., the length scales of the the radial basis functions?