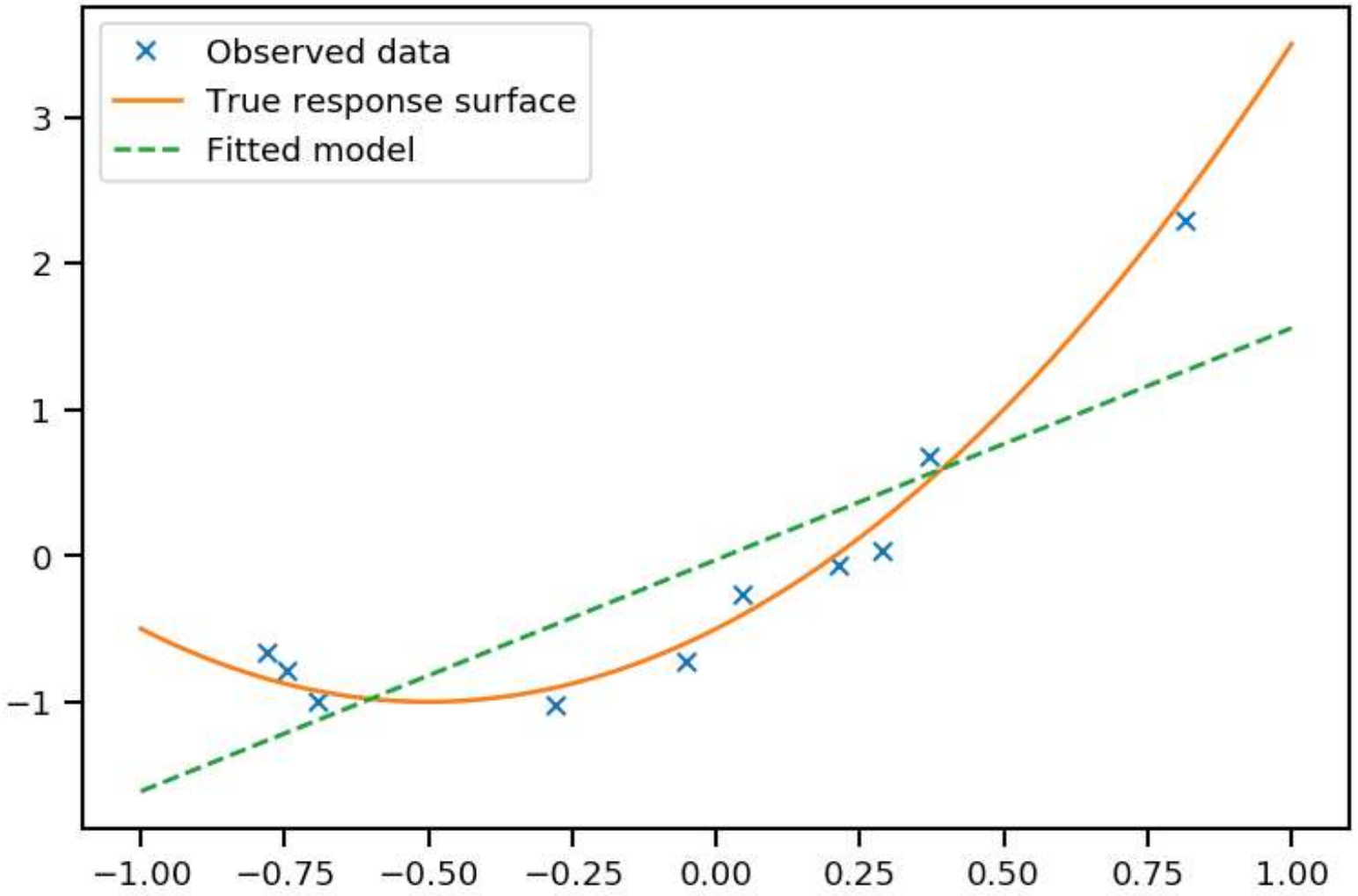
Lecture 13: Linear Regression via Least Squares

Professor Ilias Bilionis

Polynomial regression

An example that doesn't work



Regression model

$$y = w_0 + w_1 \cdot x + w_2 \cdot x^2$$

$$\underline{w} = (w_0, w_1, w_2)$$

Least squares loss function

$$L(\underline{w}) = \sum_{i=1}^{n} \left(y_i - w_0 - w_1 \cdot x_i - w_2 \cdot x_i^2\right)^2$$

$$= \| \underline{y} - \underline{\underline{X}}\, \underline{w} \|_2^2$$

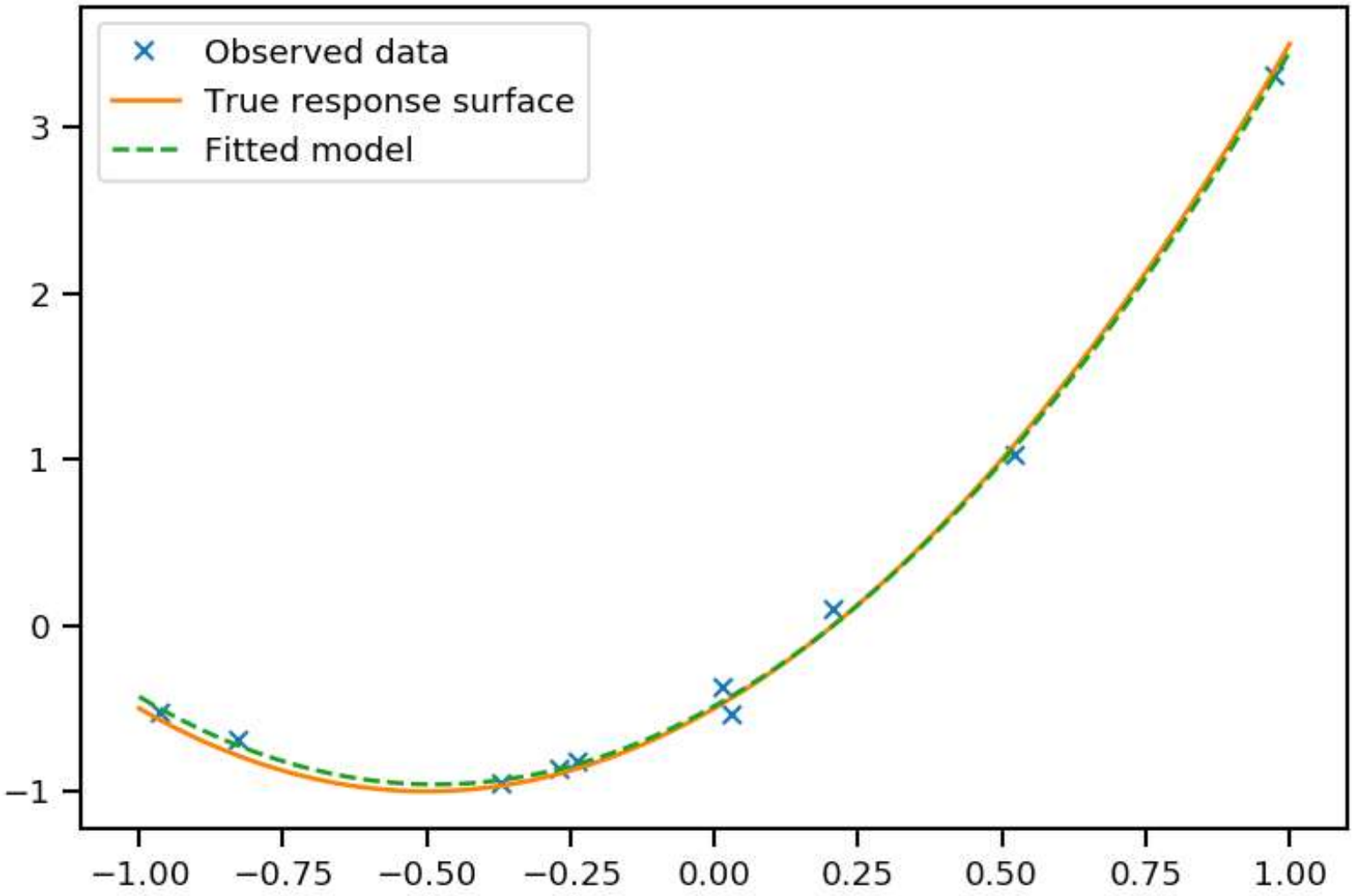
$$\underline{\underline{X}} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad (n \times \underline{3})$$

Minimizing the loss function

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

$$\nabla_{\underline{w}} L(\underline{w}) = 0 \Rightarrow \left(\underline{\underline{X}}^T \underline{\underline{X}} \right) \underline{w} = \underline{\underline{X}}^T \underline{y}$$

Example


Observed data
True response surface
Fitted model
3
2
1
0
−1
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Higher degree polynomials

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_p \cdot x^p$$

$$L(\underline{w}) = \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i - w_2 x_i^2 - \cdots - w_p x_i^p \right)^2$$

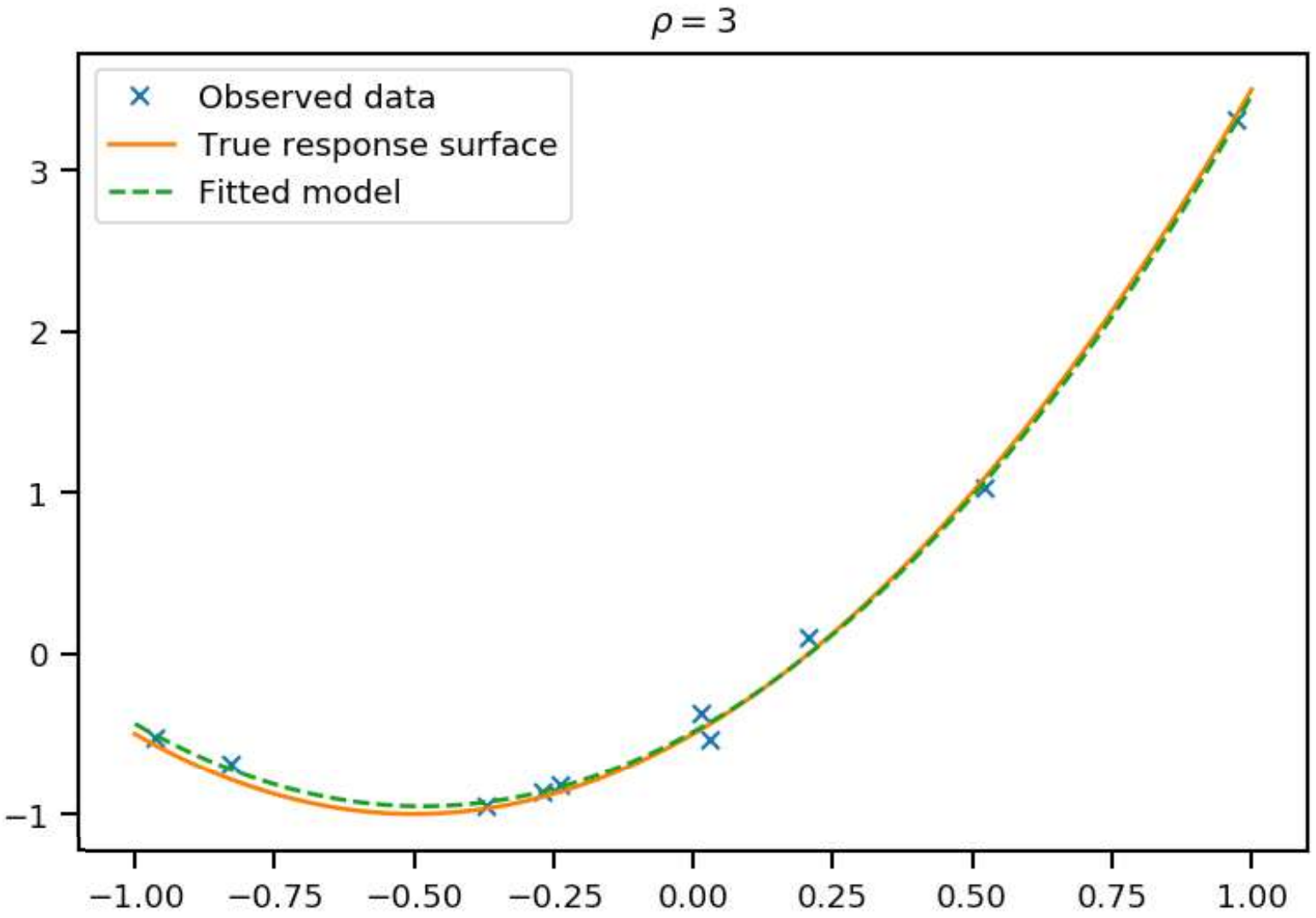
$$= \| \underline{y} - \underline{\underline{X}} \cdot \underline{w} \|_2^2$$

$$\underline{w} = (w_0, w_1, \ldots, w_p)$$

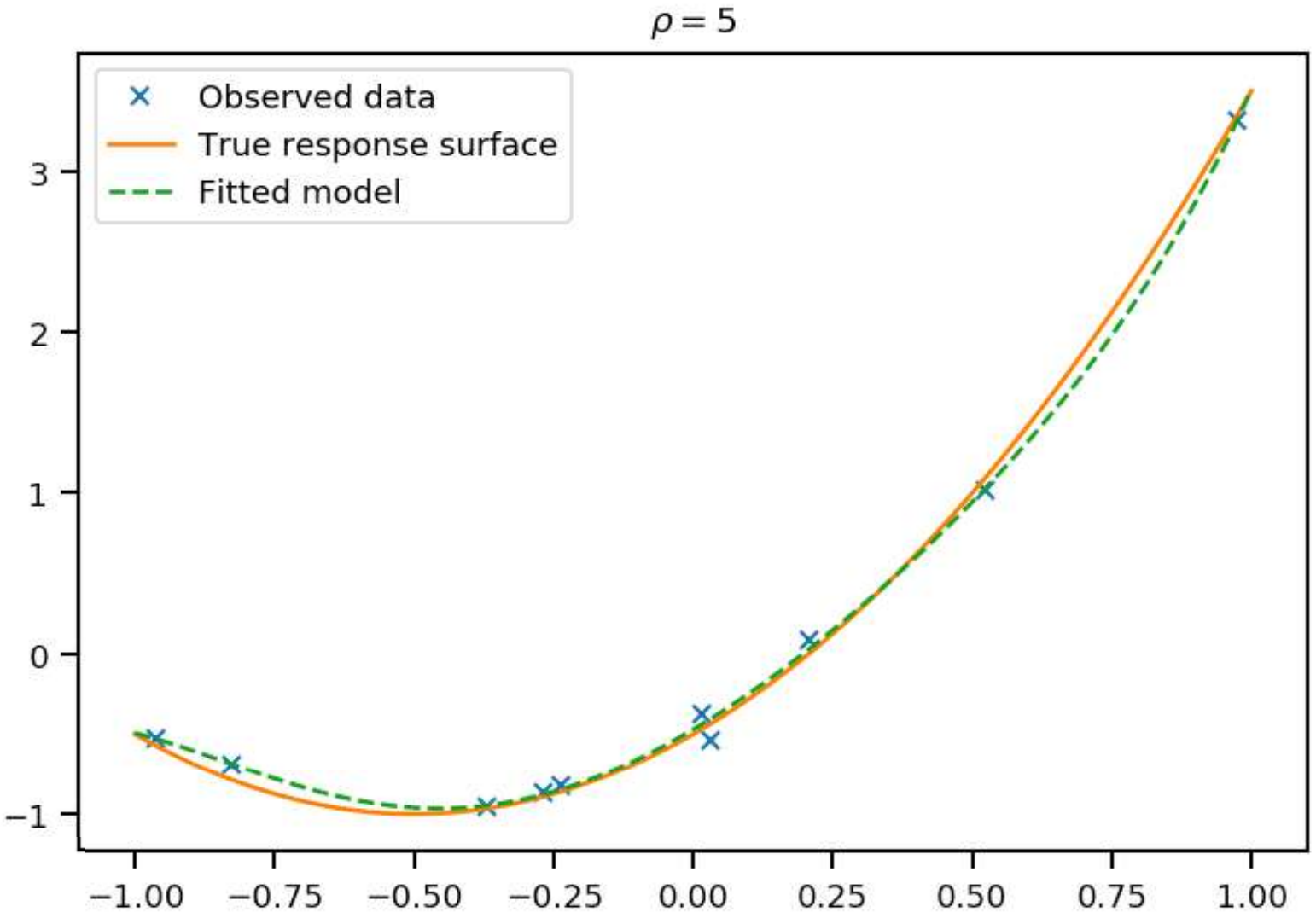
$$\underline{\underline{X}} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{pmatrix} \quad (n \times (p+1))$$

$$\nabla_{\underline{w}} L(\underline{w}) = 0 \Rightarrow \quad \underline{\underline{X}}^T \underline{\underline{X}} \cdot \underline{w} = \underline{\underline{X}}^T \cdot \underline{y}$$

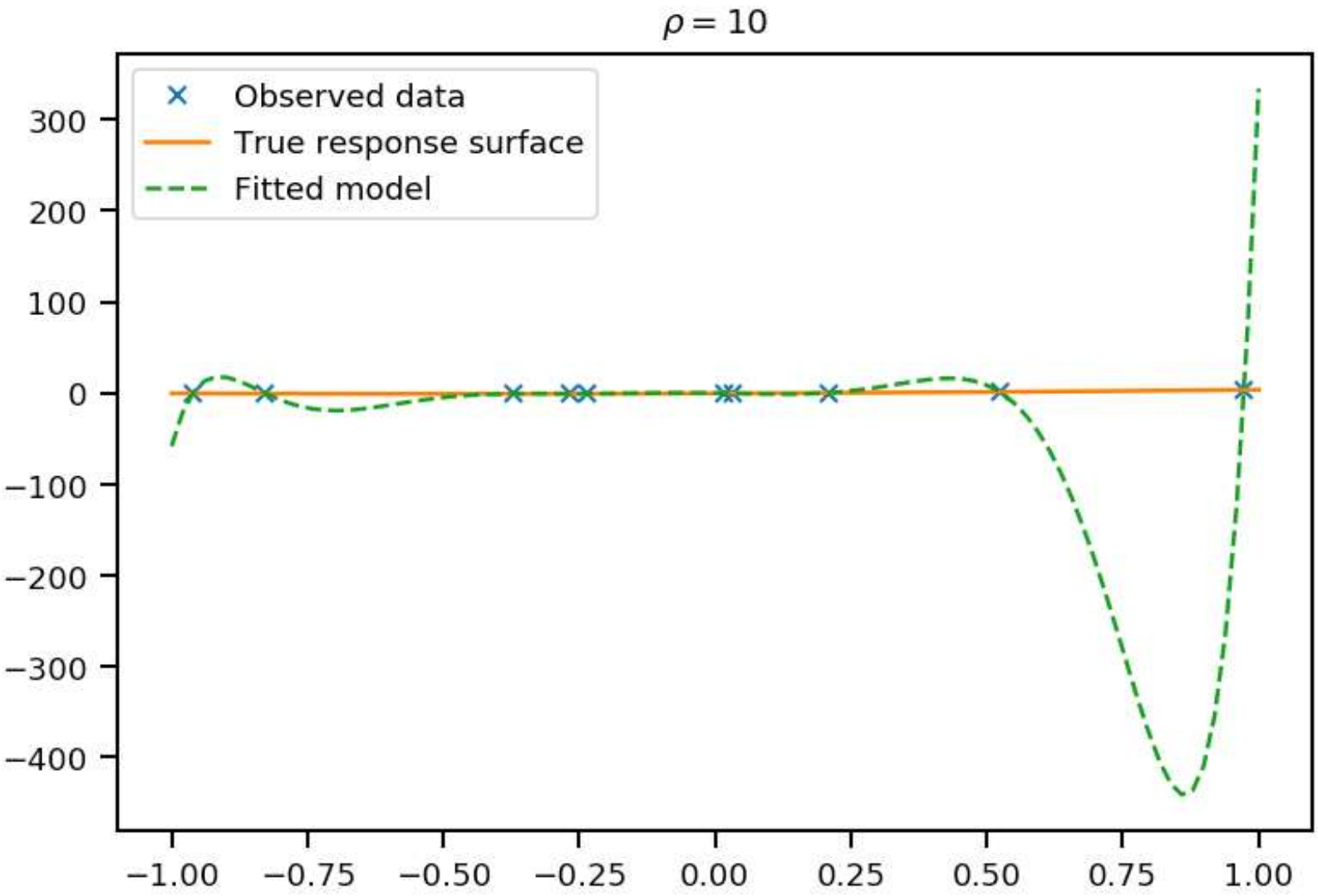
Example


$\rho = 3$
Observed data
True response surface
Fitted model
3
2
1
0
−1
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Example


$\rho = 5$
Observed data
True response surface
Fitted model
3
2
1
0
−1
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Example


$\rho = 10$
Observed data
True response surface
Fitted model
300
200
100
0
−100
−200
−300
−400
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00