# NONLINEAR FEATURE MAPS

Sometimes linear methods are not adequate. One way to create nonlinear estimators or classifiers is to first transform the data via a nonlinear feature map

$$\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}^{d'}$$

and apply a linear method to the transformed data $\Phi(x_1), \ldots, \Phi(x_n)$

$\boxed{\text{Regression}}$

In regression, the model becomes

$$f(x) = \beta^T \Phi(x) + \beta_0 \ , \quad \beta \in \mathbb{R}^{d'}$$

<u>Exercise</u> Suppose $f(x)$ is a cubic polynomial. Determine the least-squares estimate of $f$ given $(x_1, y_1), \ldots, (x_n, y_n)$.

Solution

$$f(x) = \beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0$$

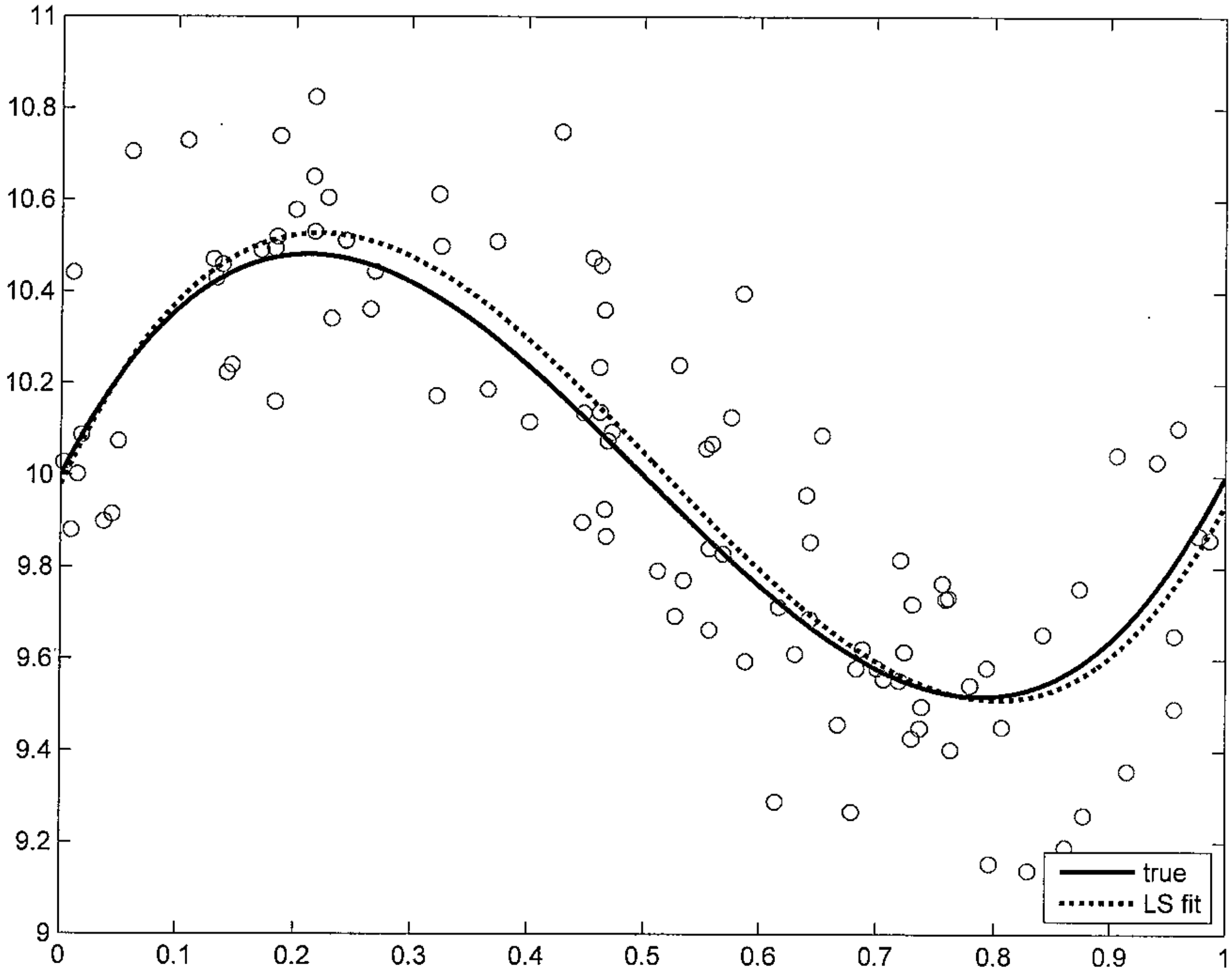
$$\Longrightarrow \quad \Phi_k(x) = x^k$$

$$\Longrightarrow \quad A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

$$\Longrightarrow \quad \hat{\theta} = (A^T A)^{-1} A^T \underline{y}$$

gives the LS cubic polynomial fit.

What if a polynomial model is also not appropriate, or the degree is unknown? We'll address these and other issues later in the course.



11
10.8
10.6
10.4
10.2
10
9.8
9.6
9.4
9.2
9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
true
LS fit

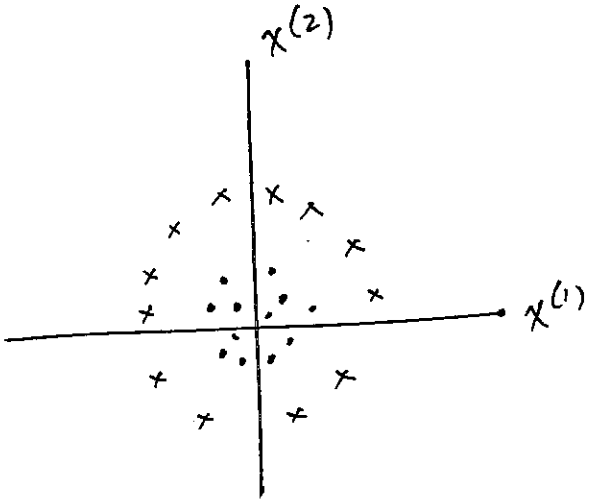
# Classification

In classification, the classifier is

$$\mathrm{sign}\{ W^T \Phi(x) + b \}$$

where $w \in \mathbb{R}^{d'}$.

## Example 1




$$x \longmapsto \Phi(x) := \begin{bmatrix} 1 \\ x^{(1)} \\ x^{(2)} \\ x^{(1)} x^{(2)} \\ (x^{(1)})^2 \\ (x^{(2)})^2 \end{bmatrix}$$

Then the data are linearly separable in the new feature space. They are correctly classified by

$$\mathrm{sign}\{ W^T \Phi(x) \}$$

where

$$W = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$