Approximation and Optimization

Problem: Approximate $f(\mathbf{x})$ by $f_a(\mathbf{x}, \mathbf{w})$.

Classical: f is known and f_a is a restricted form.
Learning: Only know a set of noisy points of f.

An approximation has the following elements

1. Form: What kind of function is f_a?
2. Norm: How to measure the error of f_a?
3. Algorithm: How to find values for $\mathbf{w}$?
   a. Efficiency: How much time is needed?
   b. Optimality: Minimize error on points?
   c. Generalization: Minimize error on f?

Univariate Forms

Polynomials: $w_0 + w_1 x + \cdots + w_n x^n$

Fourier: $w_0 + v_1 \sin(x) + w_1 \cos(x) + \cdots$
$+ v_n \sin(nx) + w_n \cos(nx)$

Linear Combination of Basis Functions:

$$\sum_{i=1}^{n} w_i \, \phi_i(x) \quad \text{or} \quad \sum_{i=1}^{n} w_i \, \phi_i(\mathbf{x})$$

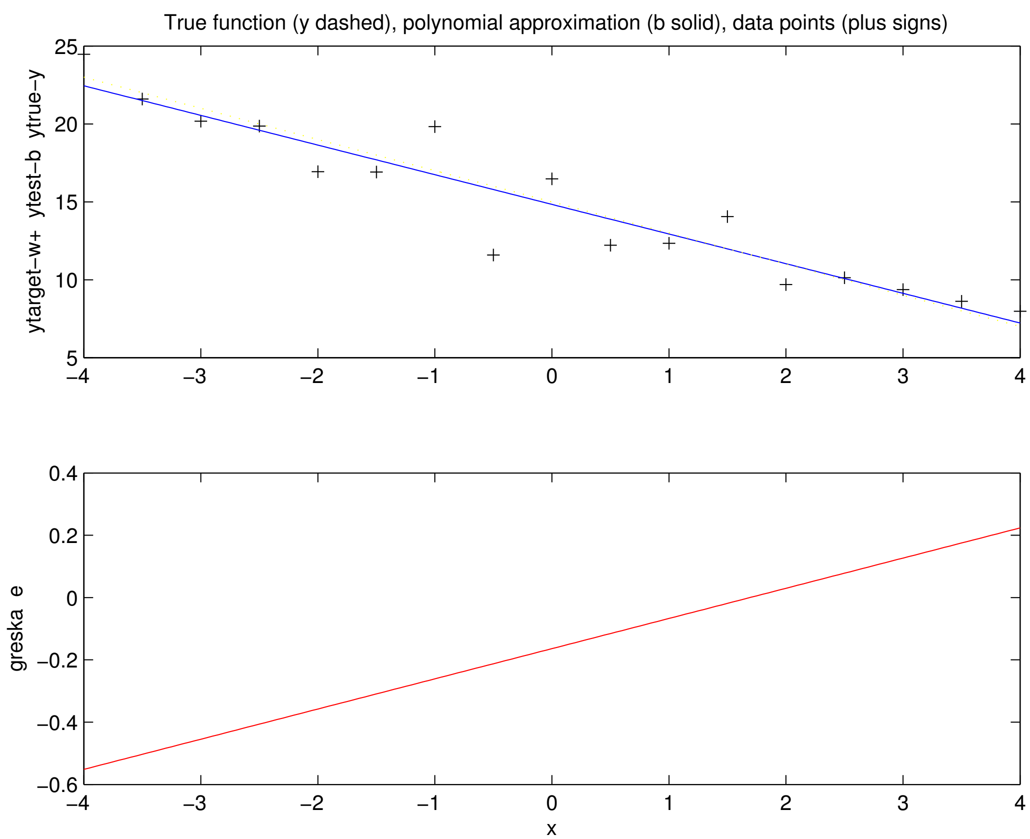
Examples: Sigmoidal, Radial BF, Splines

Example Approximations

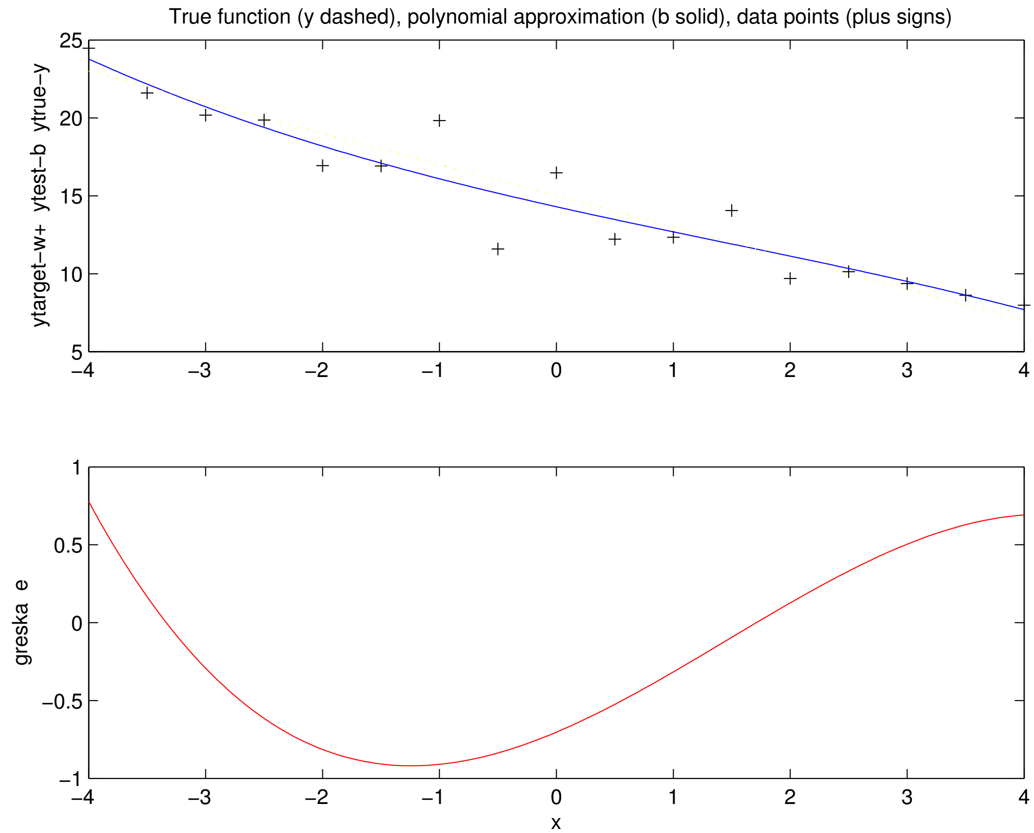
In the following figures, 17 points from a noisy $f(x) = -2x + 15$ are approximated by various polynomials.

Note that forms with more parameters will better approximate the data, but might be worse approximations of f. This is called *overfitting*.

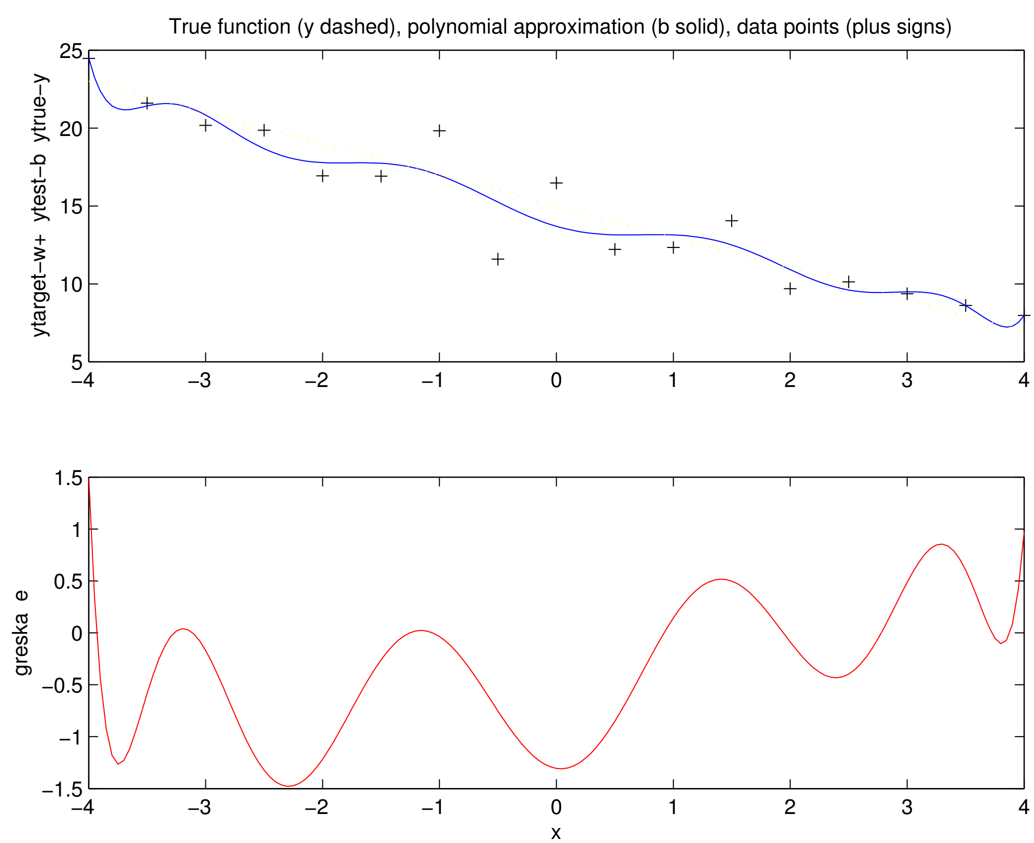
First-order polynomial



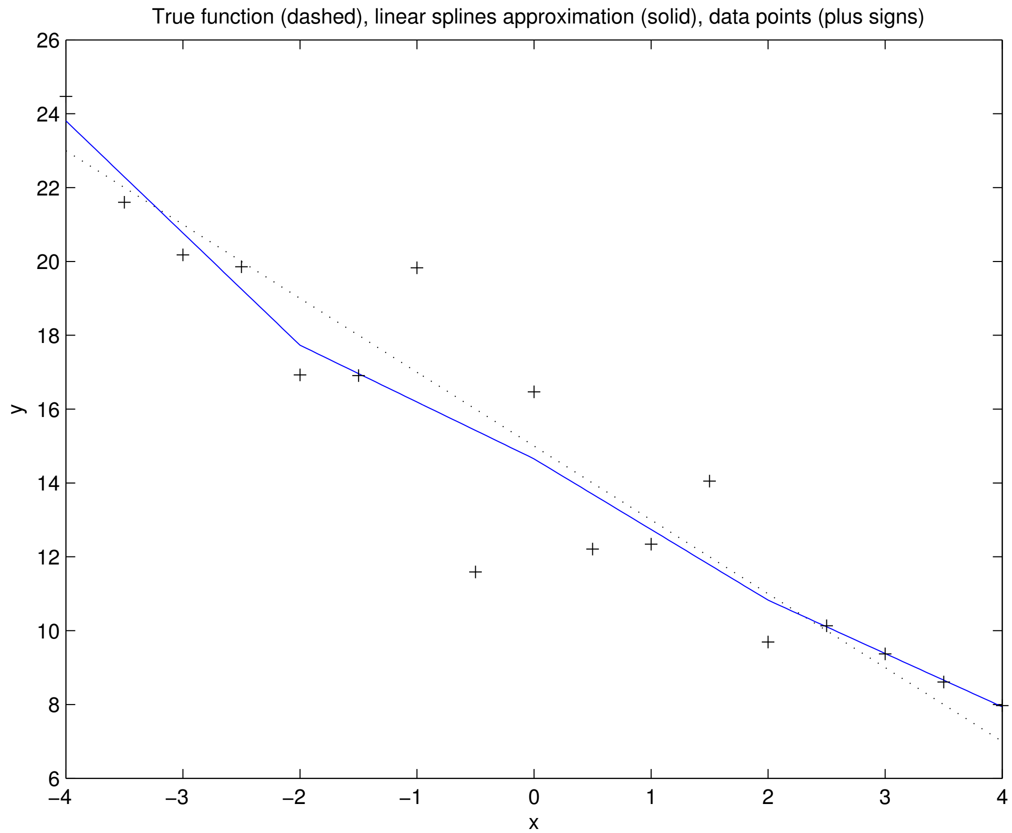
Third-order polynomial



Tenth-order polynomial



Linear splines



Multivariate Forms

Multilayer perceptrons (activation function ϕ):

$$\phi\left(b + \sum_j w_j \phi\left(a_j + \sum_i x_i v_{ij}\right)\right)$$

RBF Networks (centers $\mathbf{c}_j$):

$$\sum_j w_j \exp\left(-\frac{\Sigma_i (x_i - c_{ji})^2}{2\sigma^2}\right)$$

SVM (kernel function K, support vectors $\mathbf{x}_j$):

$$\sum_j \alpha_j K(\mathbf{x}, \mathbf{x}_j) + b$$

Error Norms

Let $\mathbf{d}$ and $\mathbf{o}$ be the desired and NN output.

L_p norm: $\|\mathbf{d} - \mathbf{o}\|_p = (\Sigma_i |d_i - o_i|^p)^{1/p}$

Squared error, Euclidean norm:

$$\|\mathbf{d} - \mathbf{o}\|_2 = \left(\Sigma_i |d_i - o_i|^2\right)^{1/2}$$

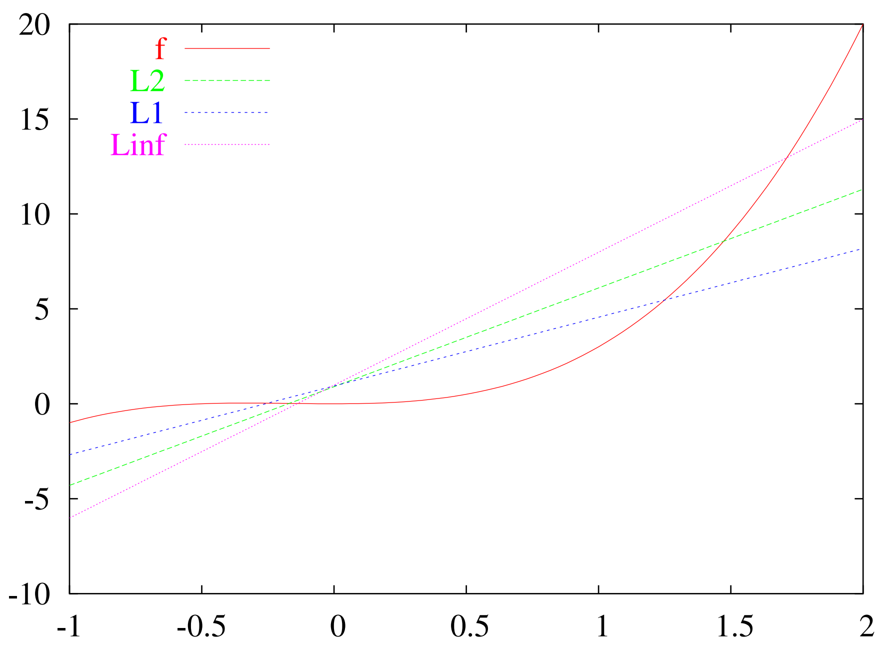
Absolute value: $\|\mathbf{d} - \mathbf{o}\|_1 = \Sigma_i |d_i - o_i|$

Chebyshev, infinity norm:

$$\|\mathbf{d} - \mathbf{o}\|_\infty = \max_i |d_i - o_i|$$

Example of Norms

This figure shows line approximatations of $f(x) = 2x^3 + x^2$ over the interval $[-1, 2]$.



Least Squares Approximation

The L_2 norm can be optimized on any data for any linear combination of basis functions, where the basis functions are fixed. This is the method of least squares.

Let $(\mathbf{x}_i, y_i)$ be the input-output pairs.
Let ϕ_j be the basis functions.
Let $\mathbf{A}$ be defined by $a_{ij} = \phi_j(x_{ij})$
Solve the set of linear equations defined by:

$$(\mathbf{A}^T\mathbf{A})\mathbf{w}^T = \mathbf{A}^T\mathbf{y}^T$$

Least squares fit of $w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2$