Supervised Learning: Regression

Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Define the set-up of Supervised Learning

Objective

Discuss basic regression models

Supervised Learning

- **The set-up:** the given training data consist of <sample, label> pairs, or (**x**, y); the objective of learning is to figure out a way to predict label y for any new sample **x**.

- Consider two types of problems:
  - **Regression**: y continuous
  - **Classification**: y is discrete, e.g., class labels.

The Task of Regression

- Given: A training set of n samples <$\mathbf{x}^{(i)}$, $y^{(i)}$> where $y^{(i)}$ is a continuous "label" (or target value) for $\mathbf{x}^{(i)}$

- To learn a model for predicting y for any new sample **x**.

- A simple model is **linear regression**: modeling the relation between y and **x** via a linear function.

$$y \approx w_0 + w_1 x_1 + \ldots + w_d x_d = \mathbf{w}^t \mathbf{x}$$

(Note: **x** is *augmented* by adding a dimension of constant 1 to the original sample.)

Linear Regression

- We can introduce an error term to capture the residual

$$y = \mathbf{w}^t\mathbf{x} + e$$

- Applying this to all *n* samples, we have :

$$\mathbf{y} = \mathrm{X}\,\mathbf{w} + \mathbf{e}$$

$$\begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix} \qquad \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_d^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_d^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1^{(n)} & x_2^{(n)} & \cdots & x_d^{(n)} \end{bmatrix} \qquad \begin{pmatrix} e^{(1)} \\ e^{(2)} \\ \vdots \\ e^{(n)} \end{pmatrix}$$

- *Learning* in this case is to figure out a good **w**.

Linear Regression (cont'd)

- Find an optimal **w** by minimizing the squared error

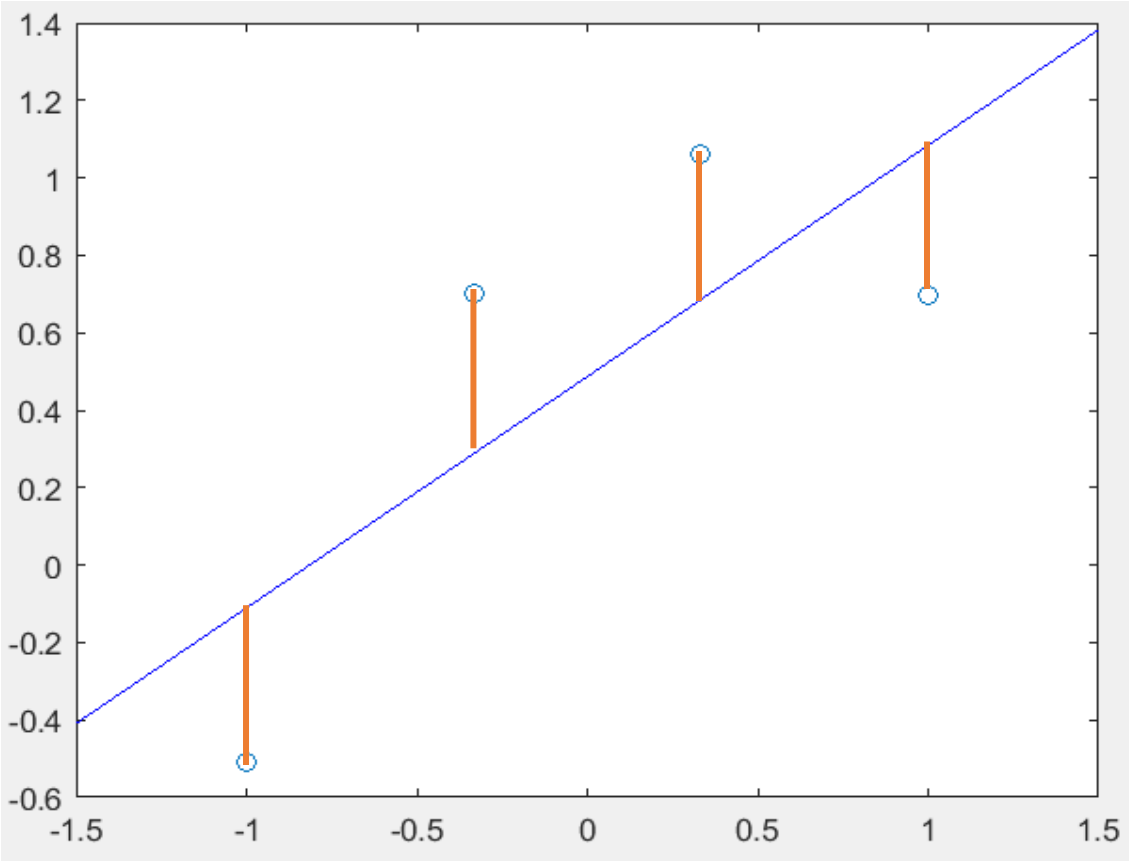
$$||\mathbf{e}||^2 = ||\mathbf{y} - \mathrm{X}\,\mathbf{w}||^2$$

- The solution can be found to be:

$$\hat{\mathbf{w}} = (\mathrm{X}^t\mathrm{X})^{-1}\mathrm{X}^t\mathbf{y}$$

- In practice, some iterative approaches may be used (e.g., gradient descent search).

A simple example



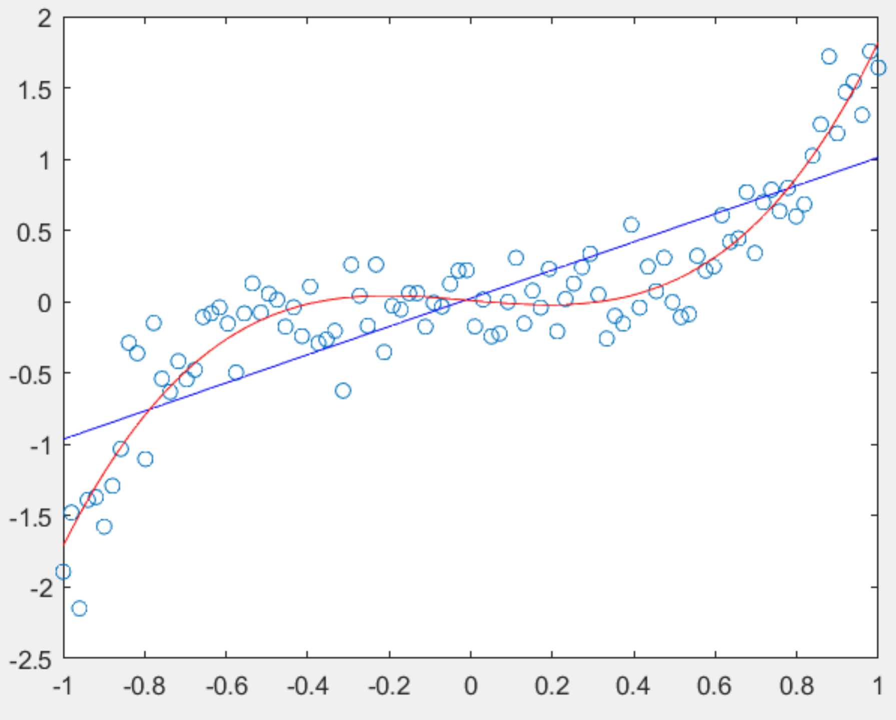
Generalizing Linear Regression

Introducing some **basis functions** $\phi_j(\mathbf{x})$:

$$y = w_0 + w_1\phi_1(\mathbf{x}) + \ \dots + w_{M-1}\phi_{M-1}(\mathbf{x})$$

Compare:

- Blue: Linear Regression
- Red: With $\phi_j(x) = x^j$



Regularized Least Squares

- E.g., use a new error function:

$$E_D(\mathbf{w}) + \lambda E_{\mathbf{W}}(\mathbf{w})$$

  - λ is the regularization coefficient
  - $E_D(\mathbf{w})$ is the data-dependent error
  - $E_{\mathbf{W}}(\mathbf{w})$is the $regularization\ term$, e.g., $\boldsymbol{E}_{\mathbf{W}}(\mathrm{w}) = ||\mathbf{w}||^{\boldsymbol{q}}$

- Help to alleviate overfitting.