Supervised Learning

Regression Example:

Suppose you receive a table of prices for houses along with their respective house sizes.

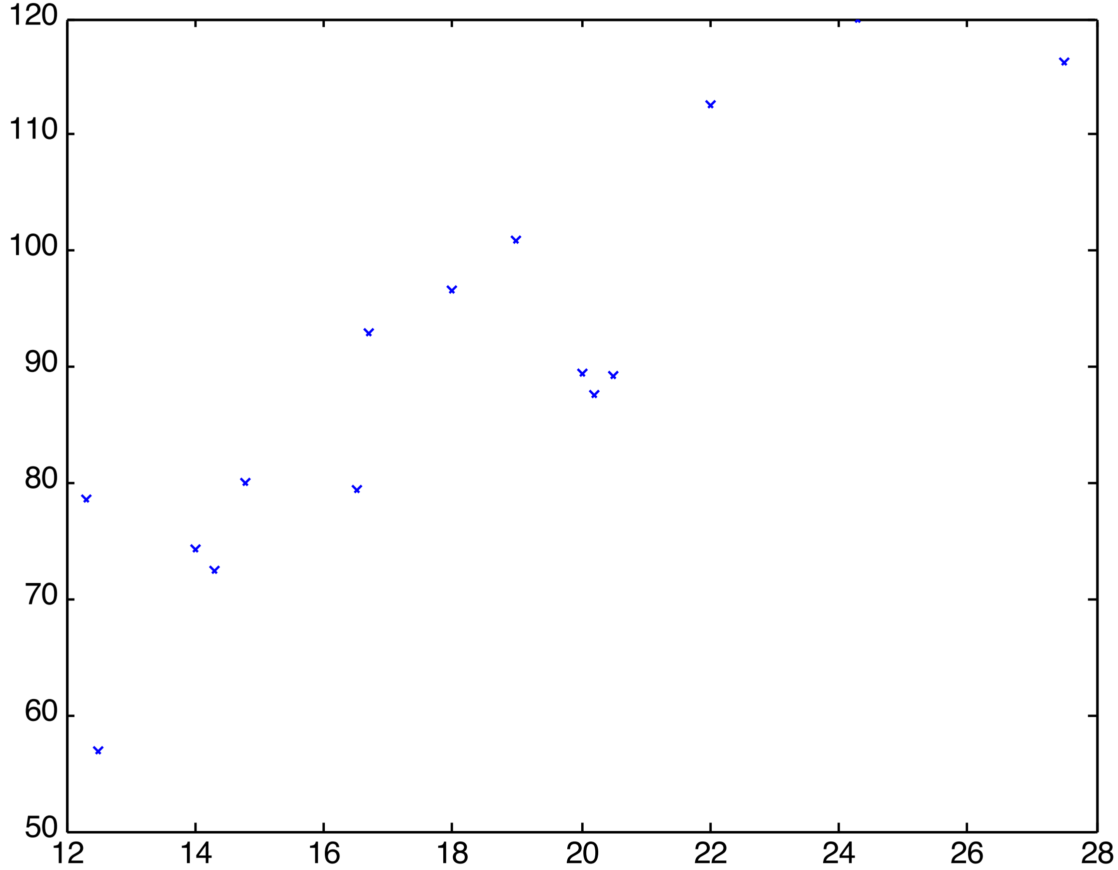
House number	Area (100 sq. feet)	Price (1000 USD)
1	20	89.5
2	14.8	79.9
3	20.5	89.1
4	12.5	56.9
5	18.0	96.6
6	14.3	72.5
7	27.5	116.3
8	16.5	79.3
9	24.3	119.9
10	20.2	87.6
11	22.0	112.6
12	19	100.8
13	12.3	78.5
14	14	74.3
15	16.7	92.8

modified from: Hung Chen, http://www.math.ntu.edu.tw/~hchen/

Supervised Learning

Regression Example:

Suppose you receive a table of prices for houses along with their respective house sizes.



Exercise:

Predict the price for a house with 2300 square feet area!

Supervised Learning

Regression:

If the target space L consists of real numbers (continuous values), then the learning task is a *regression* problem.

"Given an input value, predict the function value."

A function $f: F \rightarrow L$ is called a *regression function* in this case.

If f is a linear function (i.e., $f(x) = \theta_0 + \theta_1 x$), the problem is also called *linear regression*.

Supervised Learning

Recapitulation:

We can find the "best" parameters $\hat{\theta}_0$ and $\hat{\theta}_1$ by minimizing the squared error between the model prediction and the actual housing prices:

$$Loss(\hat{\theta}_0, \hat{\theta}_1) = \sum_{i=1}^{n} \left(y_i - f_{\hat{\theta}_0,\hat{\theta}_1}(x)\right)^2 = \sum_{i=1}^{n} (y_i - (\hat{\theta}_0 + \hat{\theta}_1 x_i))^2$$

The loss is minimized when its partial derivatives w.r.t. $\hat{\theta}_0$ and $\hat{\theta}_1$ are 0:

$$\frac{\delta}{\delta\hat{\theta}_0} \sum_{i=1}^{n} (y_i - (\hat{\theta}_0 + \hat{\theta}_1 x_i))^2 = 0 \text{ and } \frac{\delta}{\delta\hat{\theta}_1} \sum_{i=1}^{n} (y_i - (\hat{\theta}_0 + \hat{\theta}_1 x_i))^2 = 0$$

Supervised Learning

Recapitulation:

These equations have a unique solution:

$$\hat{\theta}_1 = \frac{n\sum_{j=1}^{n} x_j y_j - \left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} y_j\right)}{n\left(\sum_{j=1}^{n} {x_j}^2\right) - \left(\sum_{j=1}^{n} x_j\right)^2}$$

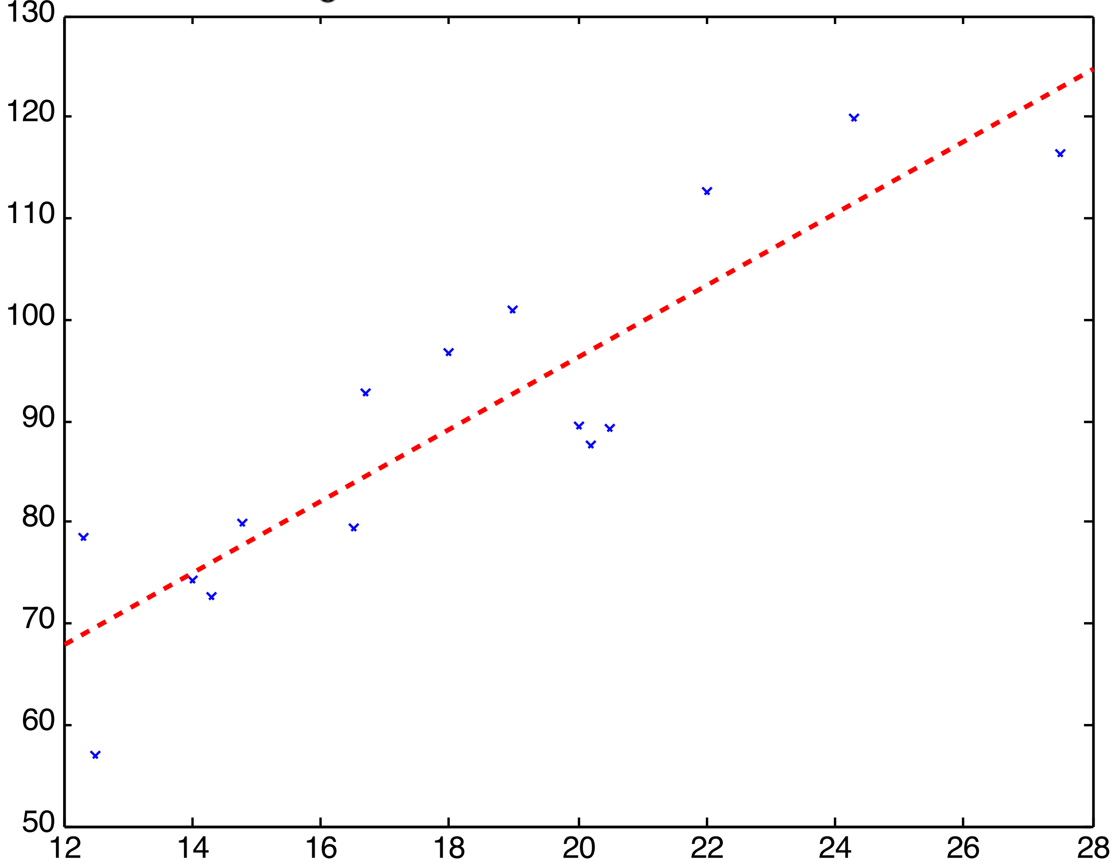
and

$$\hat{\theta}_0 = \frac{\left(\sum_{j=1}^{n} y_j\right) - \hat{\theta}_1\left(\sum_{j=1}^{n} x_j\right)}{n}$$

Supervised Learning

In our "housing" example, the method of least squares (least squares regression) leads to the optimal parameters

$\hat{\theta}_1 = 3.547$ and $\hat{\theta}_0 = 25.306$



Supervised Learning

Homework:

Use the formula for simple linear regression to verify the values of $\hat{\theta}_0$ and $\hat{\theta}_1$!

Supervised Learning

Exercise:

Suppose you're running a company, and you want to develop learning algorithms to address each of two problems.

1. You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
2. You'd like software to examine individual customer accounts, and for each account decide if it has been hacked.
3. You'd like to run a spam filter for your employees' email accounts.

Should you treat these as classification or regression problems?