Features and Polynomial Regression

Suppose we have two features, the frontage of house and the depth of the house and we might build a linear regression model where frontage is our first feature x_1 and and depth is our second feature x_2.

What we can do is actually create new features by ourself. If we want to predict the price of a house, what we might do instead is decide that what really determines the size of the house is the area or the land area that we own.

Housing prices prediction model

x_1 = frontage

x_2 = depth

$h_\theta(x) = \theta_0 + \theta_1 \times frontage + \theta_2 \times depth$

Instead of that, we can get a better model:

Area: $x_1 = frontage \times depth$

$h_\theta(x) = \theta_0 + \theta_1 \times area$

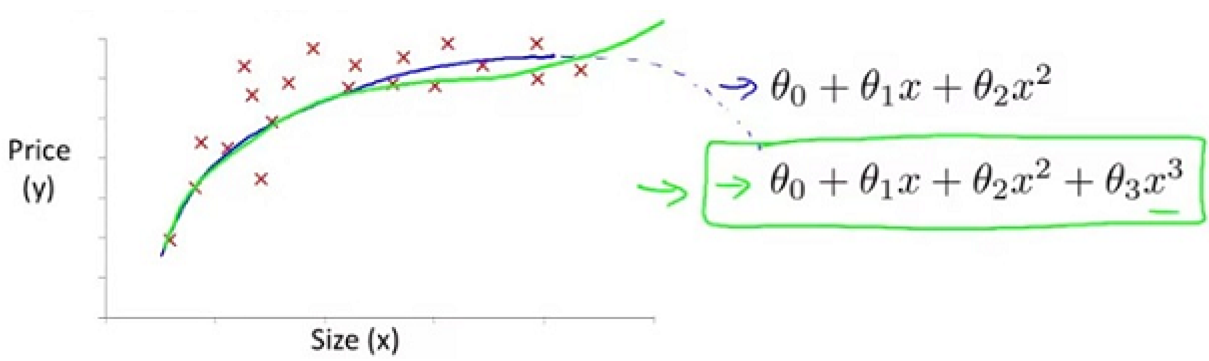
So, depending on what insight we might have into a particular problem, rather than just taking the features frontage and depth that we happen to have started off with, sometimes by defining new features we actually get a better model.

We can improve our features and the form of our hypothesis function in a couple different ways. We can **combine** multiple features into one. For example, we can combine x_1 and x_2 into a new feature x_3 by taking $x_1 \cdot x_2$.

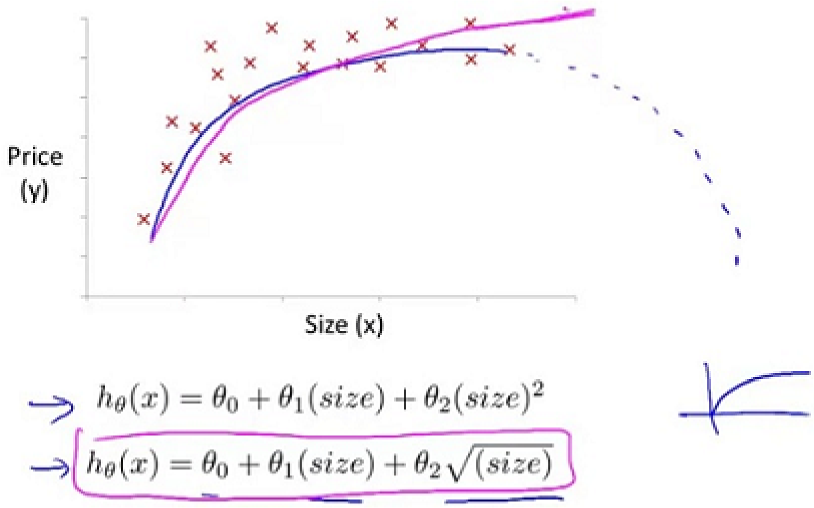
Polynomial regression

Our hypothesis function need not be linear (a straight line) if that does not fit the data well. We can **change the behavior or curve** of our hypothesis function by making it a quadratic, cubic or square root function (or any other form).

For example, if our hypothesis function is $h_\theta(x) = \theta_0 + \theta_1 x_1$ then we can create additional features based on x_1, to get the quadratic function $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$ or the cubic function $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^3$.



In the cubic version, we have created new features x_2 and x_3 where $x_2 = x_1^2$ and $x_3 = x_1^3$. To make it a square root function, we could do: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 \sqrt{x_1}$



One important thing to keep in mind is, if we choose our features this way then feature scaling becomes very important.

eg. if x_1 has range $1 - 1000$ then range of x_1^2 becomes $1 - 1000000$ and that of x_1^3 becomes $1 - 1000000000$.

Video Question: Suppose you want to predict a house's price as a function of its size. Your model is

$h_\theta(x) = \theta_0 + \theta_1(size) + \theta_2\sqrt{(size)}$.

Suppose size ranges from 1 to 1000 ($feet^2$). You will implement this by fitting a model

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

Finally, suppose you want to use feature scaling (without mean normalization).

Which of the following choices for x_1 and x_2 should you use? (Note: $\sqrt{1000} \approx 32$).

- $x_1 = size, x_2 = 32\sqrt{(size)}$
- $x_1 = 32(size), x_2 = \sqrt{(size)}$

$x_1 = \frac{size}{1000}, x_2 = \frac{\sqrt{(size)}}{32}$

- $x_1 = \frac{size}{32}, x_2 = \sqrt{(size)}$