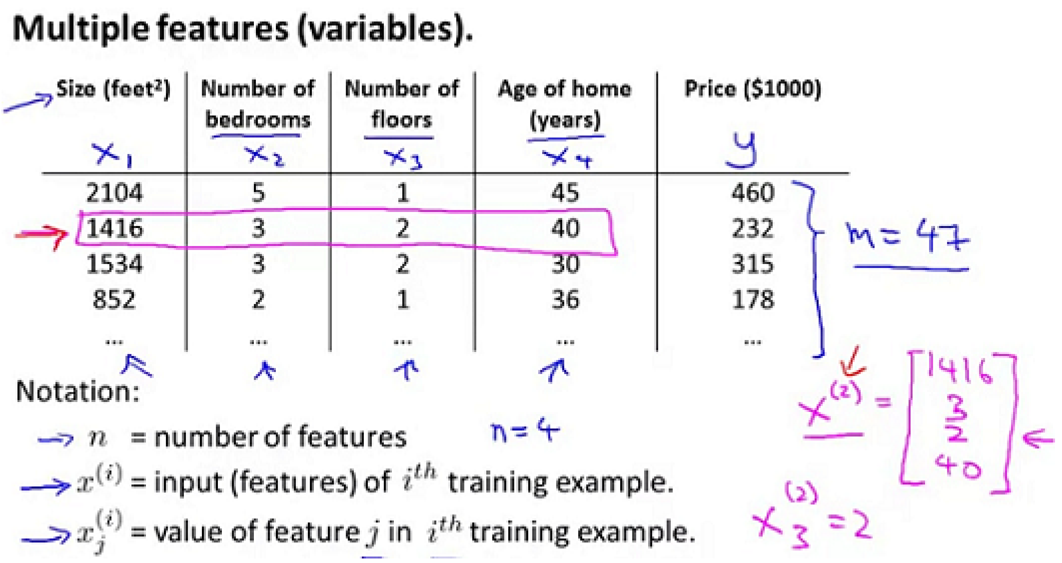
Multiple Features (Variables)

In the original version of linear regression that we developed, we have a single feature x, the size of the house, and we wanted to use that to predict y the price of the house, and that was our form of our hypothesis.



But now, we have not only the size of the house as a feature or as a variable of which to try to predict the price, but that we also know the number of bedrooms, the number of house and the age of the home and years. Now, we have more information with which to predict the price.

Features (Input variables):

- x_1 = Size ($feet^2$)
- x_2 = Number of bedrooms
- x_3 = Number of floors
- x_4 = Age of home (years)

Predictor Variables (Output variable)

- y = Price of the house ($\$1000$)

Multivariate Linear Regression works with multiple variables or with multiple features ("n" is different from our earlier notation where we were using "n" to denote the number of examples. So if we have 47 rows "m" is the number of rows on the table or the number of training examples).

Video Question:

Size ($feet^2$)	Number of bedrooms	Number of floors	Age of home (years)	Price ($\$1000$)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

In the training set above, what is $x_1^{(4)}$?

- The size (in $feet^2$) of the 1^{st} home in the training set
- The size (in years) of the 1^{st} home in the training set

> The size (in $feet^2$) of the 4^{st} home in the training set

- The size (in years) of the 4^{st} home in the training set

Now the form of the hypothesis in linear regression with multiple variables is going to be:

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$

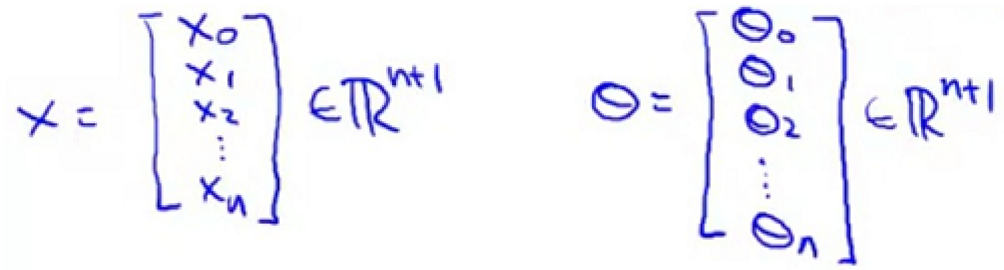
And if we have "n" features then rather than summing up over our four features, we would have a sum over our "n" features.

E.g. $h_\theta(x) = 80 + 0.1x_1 + 0.01x_2 + 3x_3 - 2x_4$

More general: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$

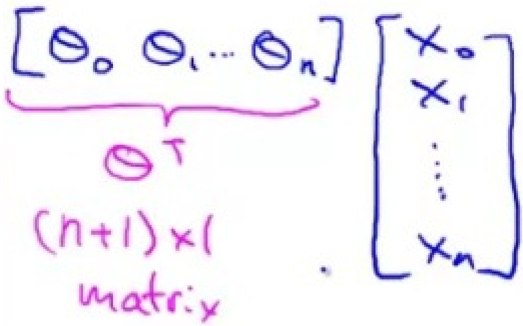
For convenience of notation, we define $x_0 = 1$. Concretely, this means that for every example "i" we have a feature vector ($x_0^i = 1$). We can think of this as defining an additional zero feature.

So now our feature vector "x" becomes to $n+1$ dimensional vector that contains is zero index, and we also going to think of our parameters as a vector and "θ" becomes to $n+1$ dimensional vector.



And now, our hypotesis function can be written as:

$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n = \theta^T x$



The notation: $h_\theta(x) = \theta^T x$ gives us a convenient way to write the form of the hypothesis as just the inner product between our parameter vector theta and our "x" vector. That's the form of our hypothesis function when we have multiple features, and the term multivariable that's just a fancy term for saying we have multiple features, or multivariables with wich to try to predict the value y.

Summary

Linear regression with multiple variables is also known as "multivariate linear regression".

We now introduce notation for equations where we can have any number of input variables.

$x_j^{(i)}$ = value of feature j in the i^{th} training example.

$x^{(i)}$ = the input (features) of the i^{th} training example.

$m =$ the number of training examples.

$n =$ the number of features.

The multivariable form of the hypothesis function accommodating these multiple features is as follows:

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots + \theta_n x_n$

In order to develop intuition about this function, we can think about θ_0 as the basic price of a house, θ_1 as the price per square meter, θ_2 as the price per floor, etc. x_1 will be the number of square meters in the house, x_2 the number of floors, etc.

Using the definition of matrix multiplication, our multivariable hypothesis function can be concisely represented as:

$$h_\theta(x) = \begin{bmatrix} \theta_0 & \theta_1 & \ldots & \theta_n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \ldots \\ x_n \end{bmatrix} = \theta^T x$$

This is a vectorization of our hypothesis function for one training example.

For convenience reasons we assume $x_0^{(i)} = 1$ for $(i \in 1, \ldots, m)$. This allows us to do matrix operations with theta and x. Hence making the two vectors 'θ' and $x^{(i)}$ match each other element-wise (that is, have the same number of elements: $n + 1$).