

Caractéristiques multiples

Let's take a look. In the original version of linear regression, you had a single feature x , the size of the house and you're able to predict y , the price of the house. The model was f_{WB} of x equals $wx + b$. But now, what if you did not only have the size of the house as a feature with which to try to predict the price, but if you also knew the number of bedrooms, the number of floors and the age of the home in years. It seems like this would give you a lot more information with which to predict the price.

Multiple features (variables)

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

To introduce a little bit of new notation, we're going to use the variables X_1 , X_2 , X_3 and X_4 , to denote the four features. For simplicity, let's introduce a little bit more notation. We'll write X subscript j or sometimes I'll just say for short, X sub j , to represent the list of features. Here, j will go from one to four, because we have four features. I'm going to use lowercase n to denote the total number of features, so in this example, n is equal to 4. As before, we'll use X superscript i to denote the i th training example. Here X superscript i is actually going to be a list of four numbers, or sometimes we'll call this a vector that includes all the features of the i th training example.

As a concrete example, X superscript in parentheses 2, will be a vector of the features for the second training example, so it will equal to this 1416, 3, 2 and 40 and technically, I'm writing these numbers in a row, so sometimes this is called a row vector rather than a column vector. But if you don't know what the difference is, don't worry about it, it's not that important for this purpose.

Multiple features (variables)

Size in feet ²	Number of bedrooms	Number of floors	Age of home in years	Price (\$) in \$1000's
X_1	X_2	X_3	X_4	
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...

$i=2$ $j=1 \dots 4$
 $x_j = j^{\text{th}}$ feature $n=4$
 $x_i = \text{number of features}$
 $x^{(i)} = \text{features of } i^{\text{th}} \text{ training example}$
 $x_j^{(i)} = \text{value of feature } j \text{ in } i^{\text{th}} \text{ training example}$
 $x^{(2)} = [1416 \ 3 \ 2 \ 40]$
 $x_3^{(2)} = 2$

To refer to a specific feature in the i th training example, I will write X superscript i , subscript j , so for example, X superscript 2 subscript 3 will be the value of the third feature, that is the number of floors in the second training example and so that's going to be equal to 2. Sometimes in order to emphasize that this X^2 is not a number but is actually a list of numbers that is a vector, we'll draw an arrow on

top of that just to visually show that is a vector and over here as well, but you don't have to draw this arrow in your notation. You can think of the arrow as an optional signifier.

They're sometimes used just to emphasize that this is a vector and not a number.

Model:

$$\text{Previously: } f_{w,b}(x) = wx + b$$

example

$$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + b$$

$$f_{w,b}(x) = 0.1 \underset{\text{size}}{\cancel{x}_1} + 4 \underset{\text{#bedrooms}}{\cancel{x}_2} + 10 \underset{\text{#floors}}{\cancel{x}_3} + -2 \underset{\text{years}}{\cancel{x}_4} + 80 \underset{\text{base price}}{b}$$

Now that we have multiple features, let's take a look at what a model would look like. Previously, this is how we defined the model, where X was a single feature, so a single number. But now with multiple features, we're going to define it differently. Instead, the model will be, $f_{w,b}$ of X equals $w_1 x_1$ plus $w_2 x_2$ plus $w_3 x_3$ plus $w_4 x_4$ plus b .

Concretely for housing price prediction, one possible model may be that we estimate the price of the house as 0.1 times X_1 , the size of the house, plus four times X_2 , the number of bedrooms, plus ten times X_3 , the number of floors, minus 2 times X_4 , the age of the house in years plus 80. Let's think a bit about how you might interpret these parameters.

If the model is trying to predict the price of the house in thousands of dollars, you can think of this b equals 80 as saying that the base price of a house starts off at maybe \$80,000, assuming it has no size, no bedrooms, no floor and no age. You can think of this 0.1 as saying that maybe for every additional square foot, the price will increase by 0.1 \$1,000 or by \$100, because we're saying that for each square foot, the price increases by 0.1, times \$1,000, which is \$100.

Maybe for each additional bathroom, the price increases by \$4,000 and for each additional floor the price may increase by \$10,000 and for each additional year of the house's age, the price may decrease by \$2,000, because the parameter is negative 2. In general, if you have n features, then the model will look like this.

$$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$$

Here again is the definition of the model with n features. What we're going to do next is introduce a little bit of notation to rewrite this expression in a simpler but equivalent way. Let's define W as a list of numbers that list the parameters $w_1, w_2, w_3, \dots, w_n$. In mathematics, this is called a vector and sometimes to designate that this is a vector, which just means a list of numbers, I'm going to draw a little arrow on top.

You don't always have to draw this arrow and you can do so or not in your own notation, so you can think of this little arrow as just an optional signifier to remind us that this is a vector. If you've taken the linear algebra class before, you might recognize that this is a row vector as opposed to a column vector. But if you don't know what those terms mean, you don't need to worry about it.

Here again is the definition of the model with n features. What we're going to do next is introduce a little bit of notation to rewrite this expression in a simpler but equivalent way. Let's define \vec{W} as a list of numbers that list the parameters W_1, W_2, W_3 , all the way through W_n . In mathematics, this is called a vector and sometimes to designate that this is a vector, which just means a list of numbers, I'm going to draw a little arrow on top.

You don't always have to draw this arrow and you can do so or not in your own notation, so you can think of this little arrow as just an optional signifier to remind us that this is a vector. If you've taken the linear algebra class before, you might recognize that this is a row vector as opposed to a column vector. But if you don't know what those terms mean, you don't need to worry about it.

$$f_{\vec{w}, b}(\vec{x}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$$

$\vec{w} = [w_1 \ w_2 \ w_3 \ \dots \ w_n]$ parameters of the model

b is a number

vector $\vec{x} = [x_1 \ x_2 \ x_3 \ \dots \ x_n]$

$$f_{\vec{w}, b}(\vec{x}) = \vec{w} \cdot \vec{x} + b = \underset{\text{dot product}}{w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_n x_n} + b$$

multiple linear regression

The name for this type of linear regression model with multiple input features is multiple linear regression. This is in contrast to univariate regression, which has just one feature. By the way, you might think this algorithm is called multivariate regression, but that term actually refers to something else that we won't be using here.

I'm going to refer to this model as multiple linear regression. That's it for linear regression with multiple features, which is also called multiple linear regression. In order to implement this, there's a really neat trick called vectorization, which will make it much simpler to implement this and many other learning algorithms.