

Ingénierie des fonctionnalités

The choice of features can have a huge impact on your learning algorithm's performance. In fact, for many practical applications, choosing or entering the right features is a critical step to making the algorithm work well. In this video, let's take a look at how you can choose or engineer the most appropriate features for your learning algorithm. Let's take a look at feature engineering by revisiting the example of predicting the price of a house.

Say you have two features for each house. x_1 is the width of the lot size of the plots of land that the house is built on. This in real state is also called the frontage of the lot, and the second feature, x_2 , is the depth of the lot size of, let's assume the rectangular plot of land that the house was built on. Given these two features, x_1 and x_2 , you might build a model like this where f of x is w_1x_1 plus w_2x_2 plus b , where x_1 is the frontage or width, and x_2 is the depth.

Feature engineering

$$f_{\vec{w},b}(\vec{x}) = w_1 \underbrace{x_1}_{\text{frontage}} + w_2 \underbrace{x_2}_{\text{depth}} + b$$



This model might work okay. But here's another option for how you might choose a different way to use these features in the model that could be even more effective. You might notice that the area of the land can be calculated as the frontage or width times the depth. You may have an intuition that the area of the land is more predictive of the price, than the frontage and depth as separate features. You might define a new feature, x_3 , as x_1 times x_2 . This new feature x_3 is equal to the area of the plot of land.

With this feature, you can then have a model $f_{\vec{w},b}$ of x equals w_1x_1 plus w_2x_2 plus w_3x_3 plus b so that the model can now choose parameters w_1 , w_2 , and w_3 , depending on whether the data shows that the frontage or the depth or the area x_3 of the lot turns out to be the most important thing for predicting the price of the house.

What we just did, creating a new feature is an example of what's called feature engineering, in which you might use your knowledge or intuition about the problem to design new features usually by transforming or combining the original features of the problem in order to make it easier for the learning algorithm to make accurate predictions.

Depending on what insights you may have into the application, rather than just taking the features that you happen to have started off with sometimes by defining new features, you might be able to get a much better model. That's feature engineering.

Feature engineering

$$f_{\bar{w},b}(\bar{x}) = w_1 \underbrace{x_1}_{\text{frontage}} + w_2 \underbrace{x_2}_{\text{depth}} + b$$

$$\text{area} = \text{frontage} \times \text{depth}$$

$$x_3 = x_1 x_2$$

new feature

$$f_{\bar{w},b}(\bar{x}) = \underbrace{w_1}_{\text{frontage}} x_1 + \underbrace{w_2}_{\text{depth}} x_2 + \underbrace{w_3}_{\text{area}} x_3 + b$$



Feature engineering:
Using **intuition** to design
new features, by
transforming or combining
original features.