(https://google.com/racialequity?authuser=0)

Feature Crosses: Encoding Nonlinearity

ated Time: 7 minutes

In Figures 1 and 2, imagine the following:

- The blue dots represent sick trees.
- The orange dots represent healthy trees.

Figure 1. Is this a linear problem?

Can you draw a line that neatly separates the sick trees from the healthy trees? Sure. This is a linear problem. The line won't be perfect. A sick tree or two might be on the "healthy" side, but your line will be a good predictor.

Now look at the following figure:

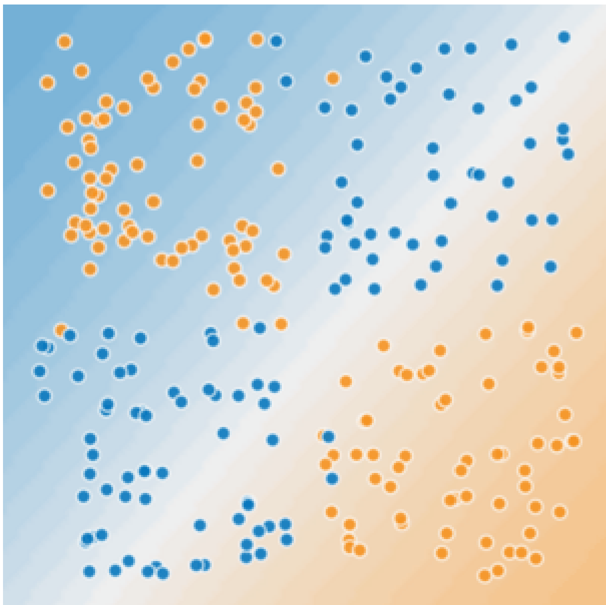

Figure 2. Is this a linear problem?

Can you draw a single straight line that neatly separates the sick trees from the healthy trees? No, you can't. This is a nonlinear problem. Any line you draw will be a poor predictor of tree health.

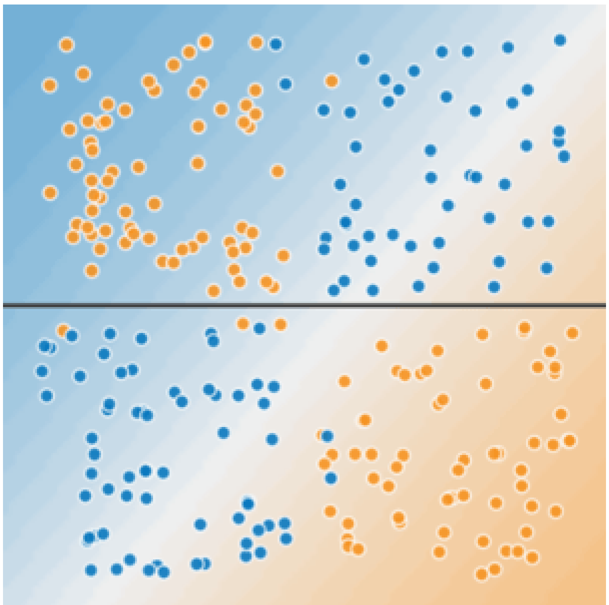

Figure 3. A single line can't separate the two classes.

To solve the nonlinear problem shown in Figure 2, create a feature cross. A **feature cross** is a synthetic feature that encodes nonlinearity in the feature space by multiplying two or more input features together. (The term *cross* comes from *cross product* (https://wikipedia.org/wiki/Cross_product).) Let's create a feature cross named x_3 by crossing x_1 and x_2:

$$x_3 = x_1 x_2$$

We treat this newly minted x_3 feature cross just like any other feature. The linear formula becomes:

$$y = b + w_1 x_1 + w_2 x_2 + w_3 x_3$$

A linear algorithm can learn a weight for w_3 just as it would for w_1 and w_2. In other words, although w_3 encodes nonlinear information, you don't need to change how the linear model trains to determine the value of w_3.

Kinds of feature crosses

We can create many different kinds of feature crosses. For example:

- `[A X B]`: a feature cross formed by multiplying the values of two features.
- `[A x B x C x D x E]`: a feature cross formed by multiplying the values of five features.
- `[A x A]`: a feature cross formed by squaring a single feature.

Thanks to stochastic gradient descent (https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent?authuser=0)
, linear models can be trained efficiently. Consequently, supplementing scaled linear models with feature crosses has traditionally been an efficient way to train on massive-scale data sets.

rms

ature crosses (https://developers.google.com/machine-learning/glossary?authuser=0#feature_cross)
nthetic feature (https://developers.google.com/machine-learning/glossary?authuser=0#synthetic_featu

Help Center (https://support.google.com/machinelearningeducation?authuser=0)

Previous
← Video Lecture
(https://developers.google.com/machine-learning/crash-course/feature-crosses/video-lecture?authuser=0)

Next