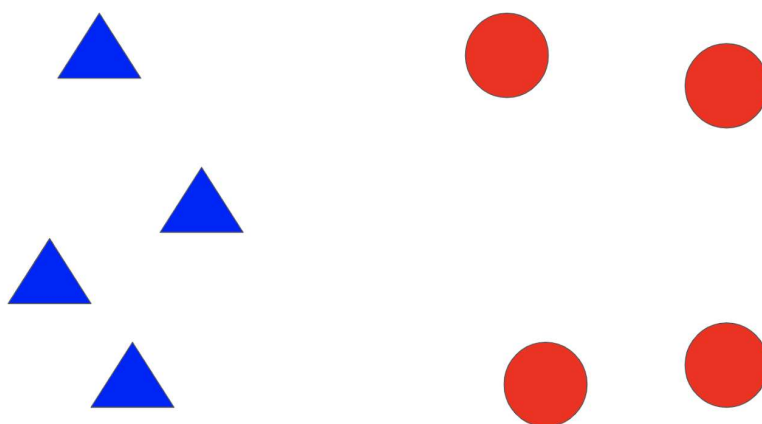# Module 3-2: Support Vector Machine

A **support vector machine** or **SVM** for short, is a popular classification algorithm. Here's how it works: say that we have data that contains two features and we aim to predict if a data point is red or blue. This is what our data might look like plotted out in a 2D space before training a classifier.


The position of the data is based on its features. For example, if each shape was a mouse then it could be plotted based on its height and weight. The color would represent the output, such as if a mouse is obese or not.

What a support vector machine aims to do is to create the best line to separate the data so that everything on the left will be blue and everything on the right will be red. But how do we know what the best line is? Is this the best line?
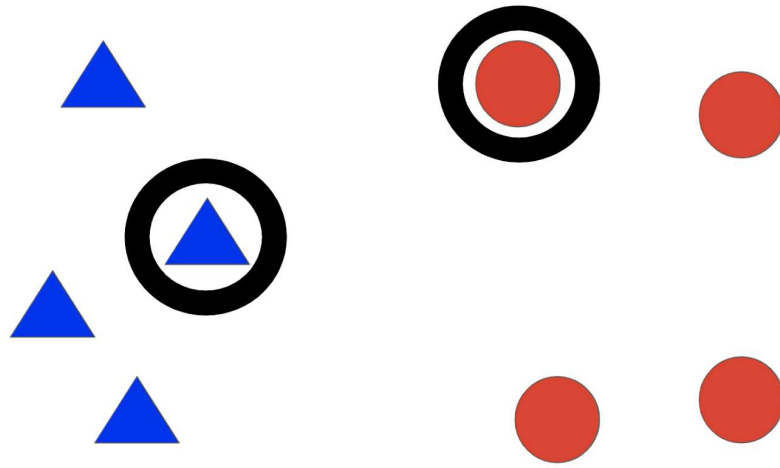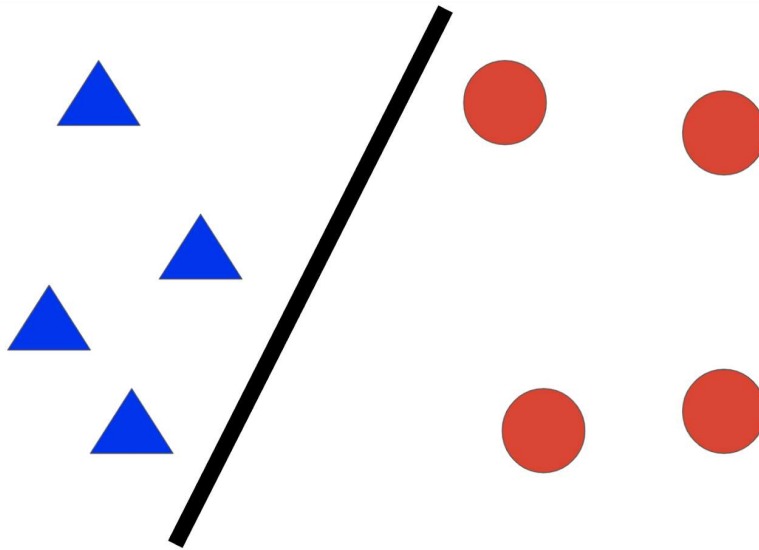
Send Issue
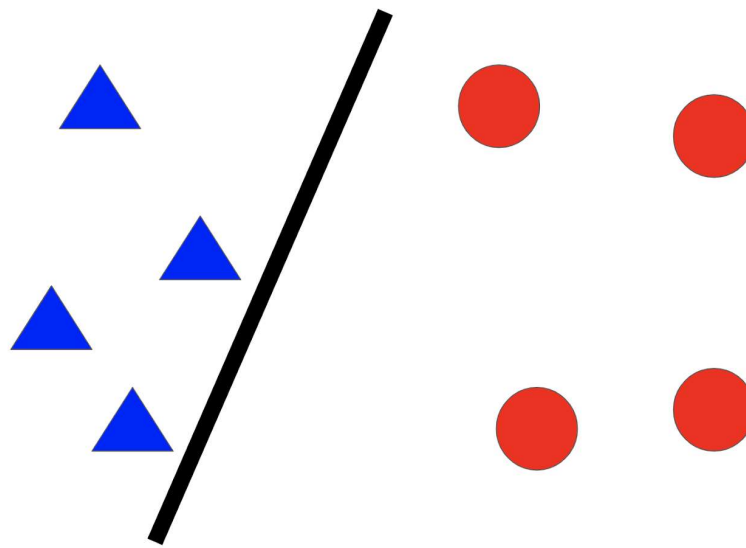
Or is this the best line?
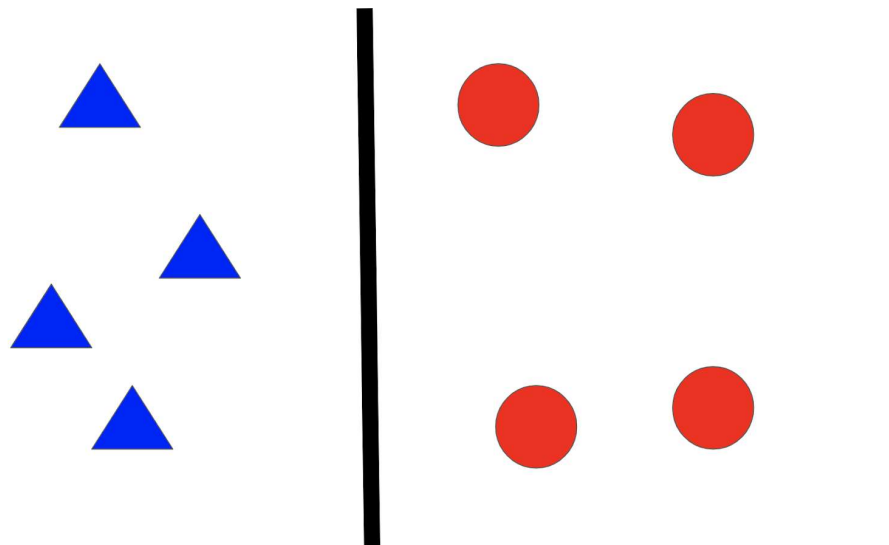
What SVM does to find the best line is draw a line between the most innermost data points from opposing classes. The two most innermost data points would be the two that have been circled in black.

The best line, also known as the decision boundary, would look something like this.

From here, we can predict that if a data point ends up on the left of the decision boundary it will be blue and if a decision boundary