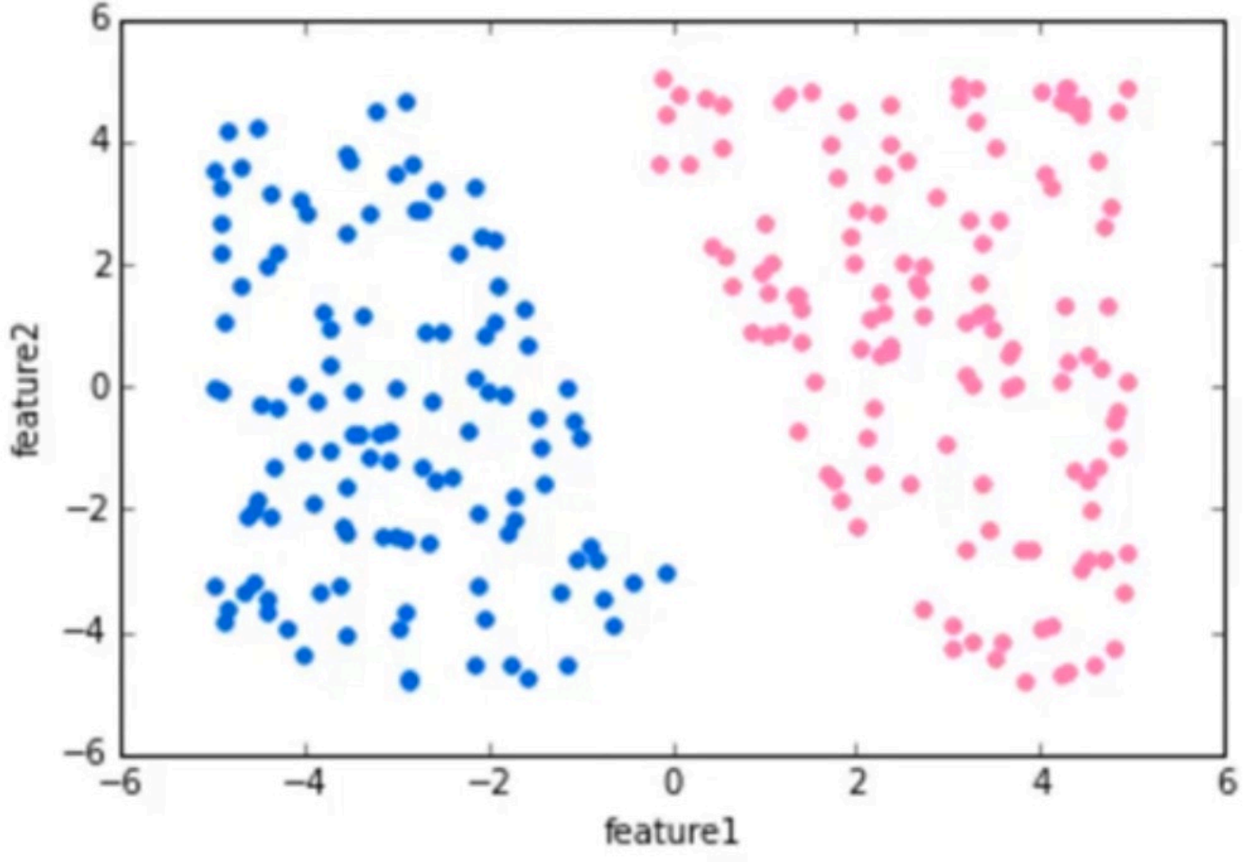
Support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

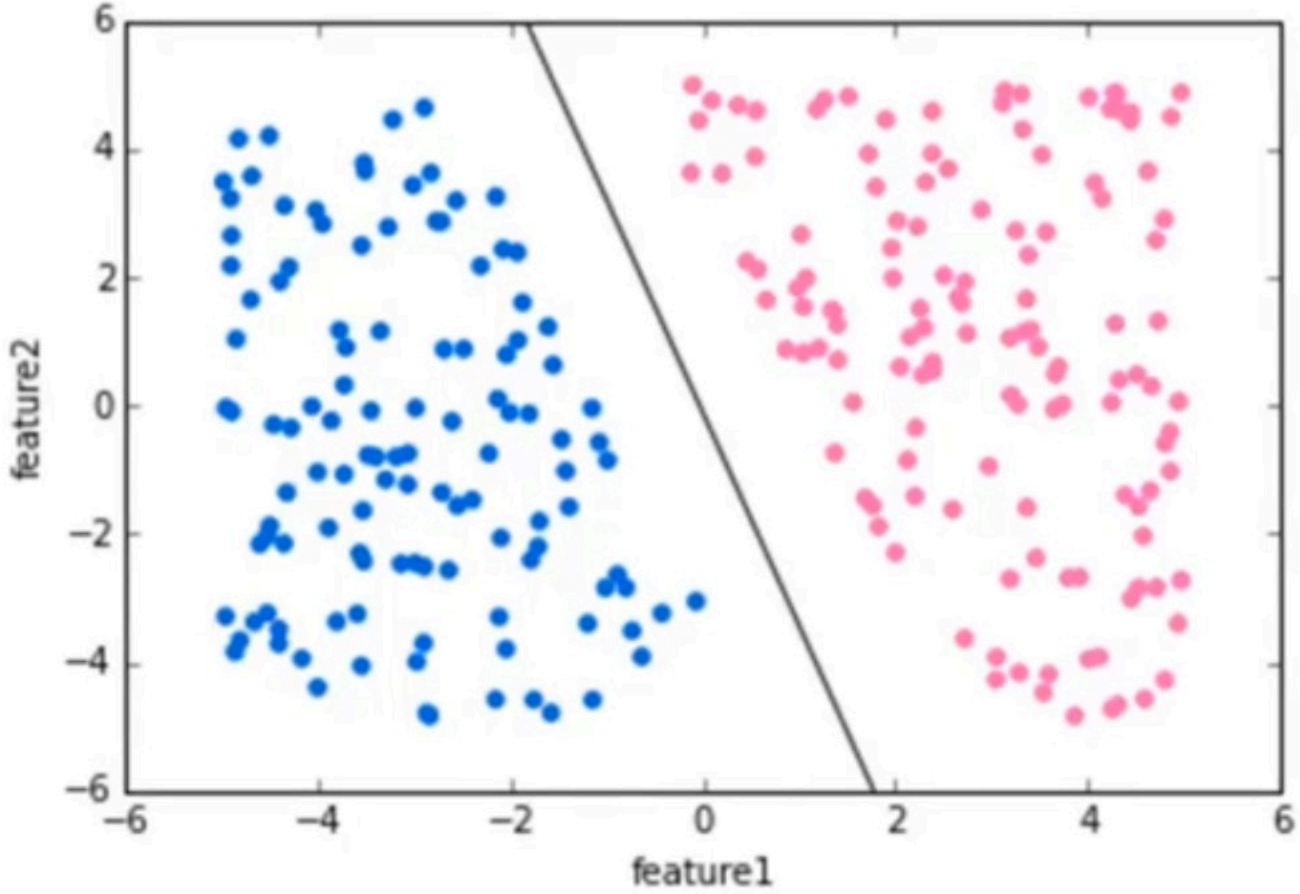
An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

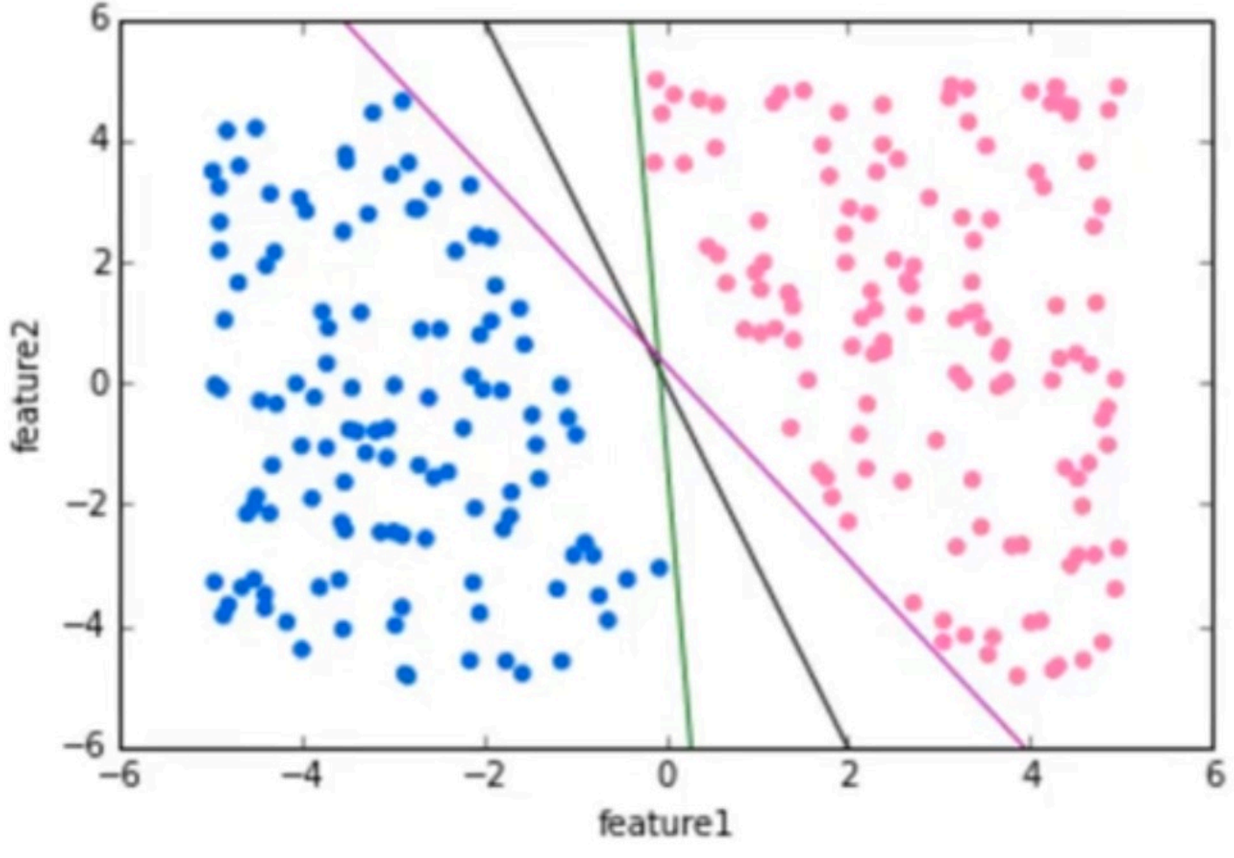
Let’s show the basic intuition behind SVMs. Imagine the labeled training data below:



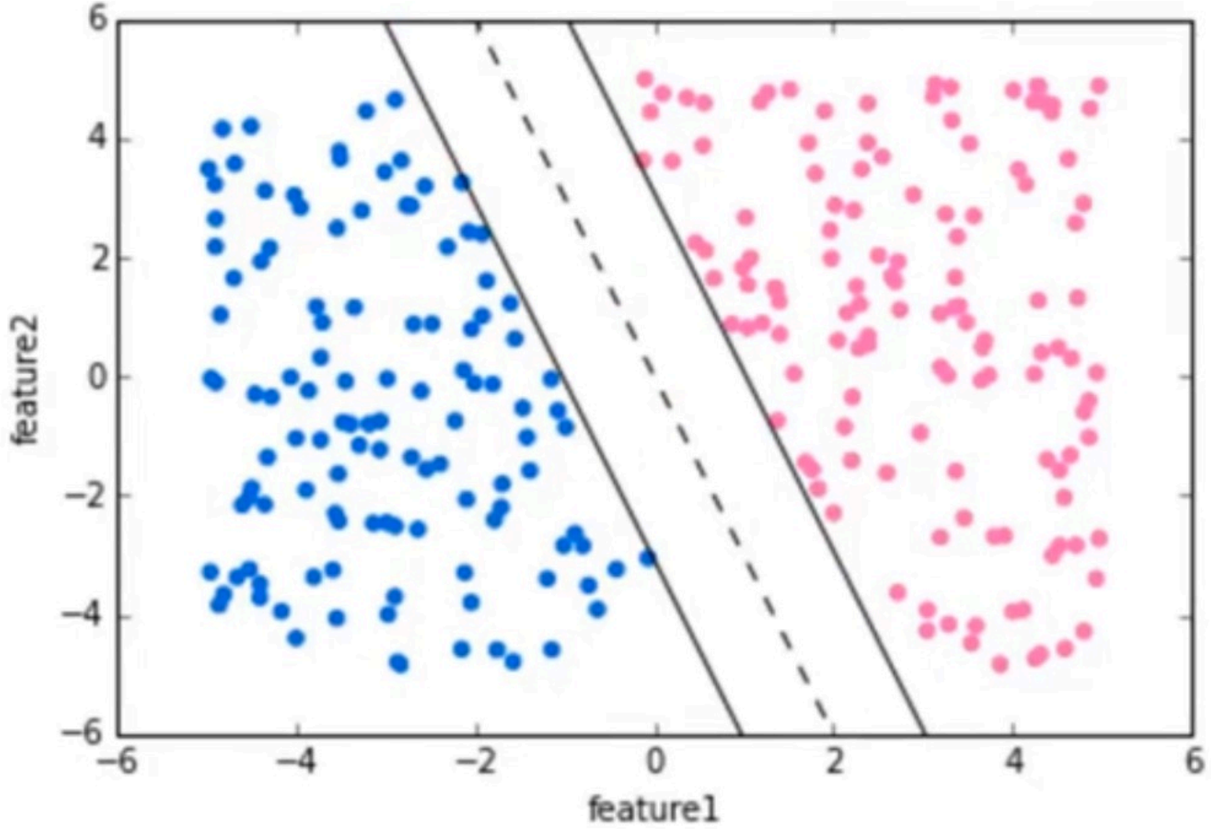
We can draw a separating “hyperplane” between the classes.



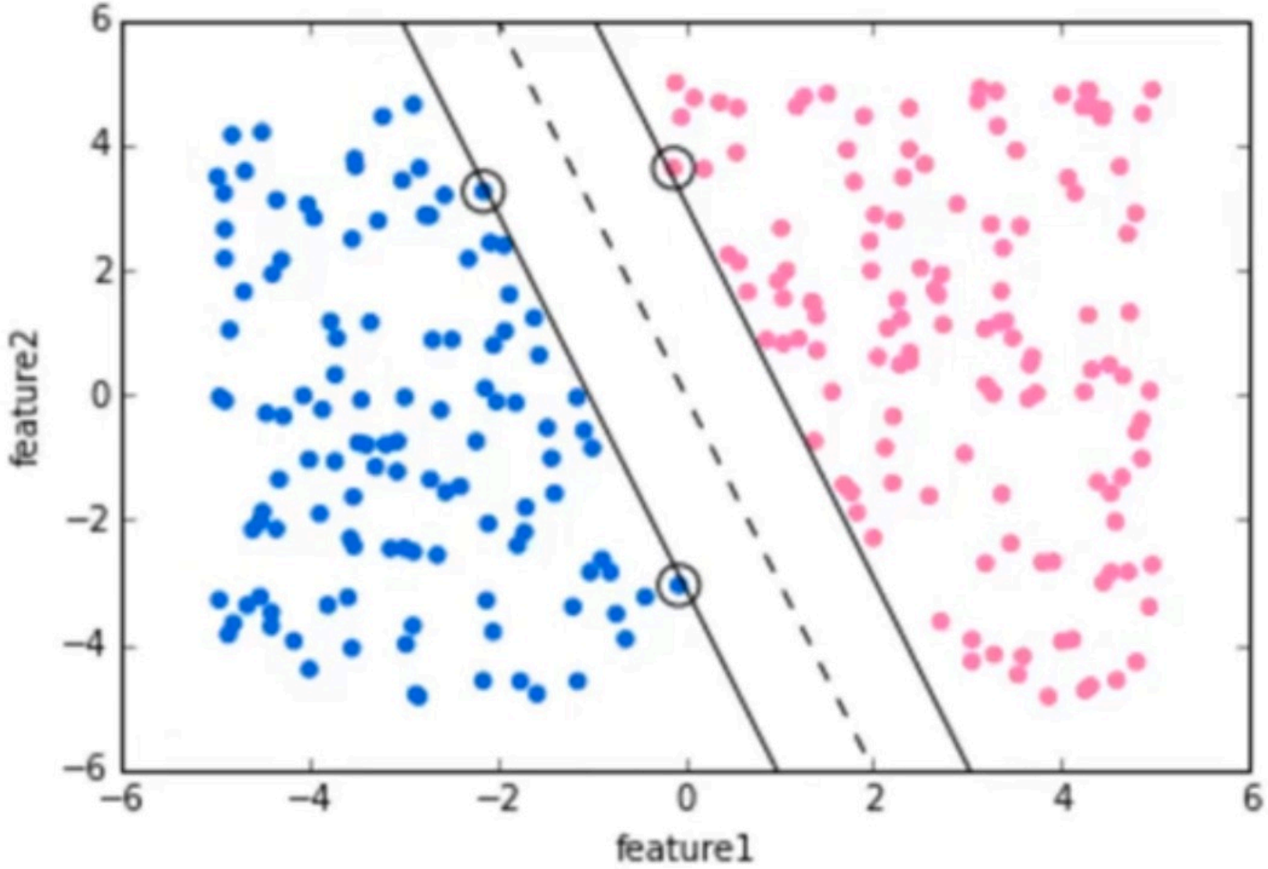
But we have many options of hyperplanes that separate perfectly...



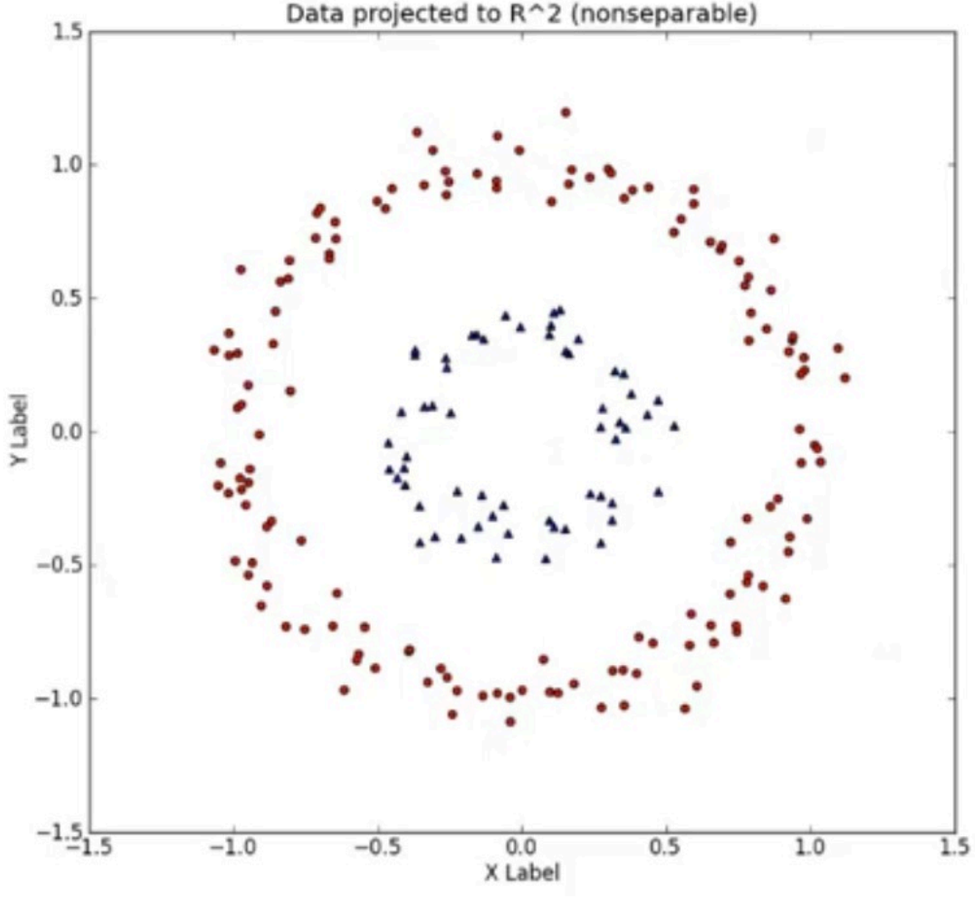
We would like to choose a hyperplane that maximizes the margin between classes

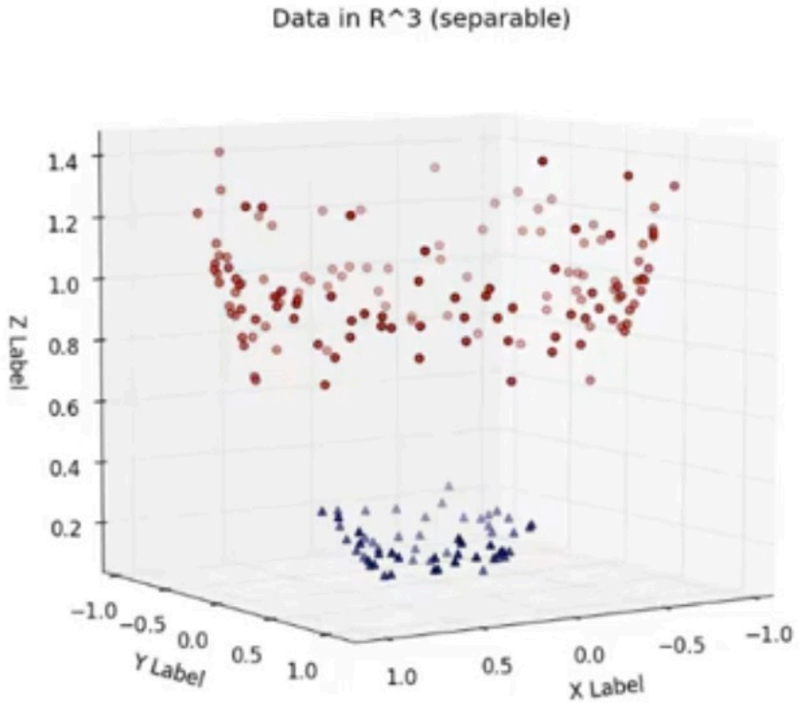


The vector points that the margin lines touch are known as Support Vectors.

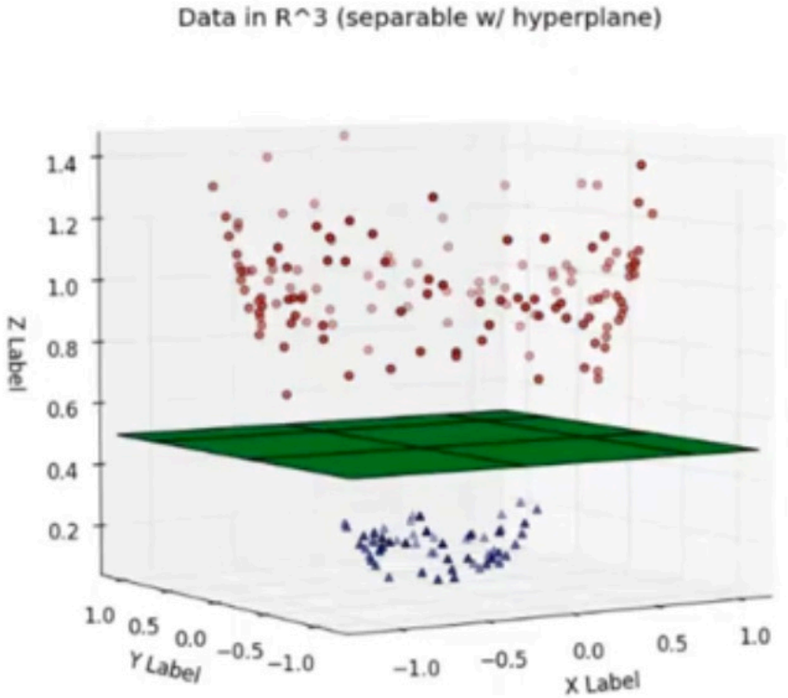


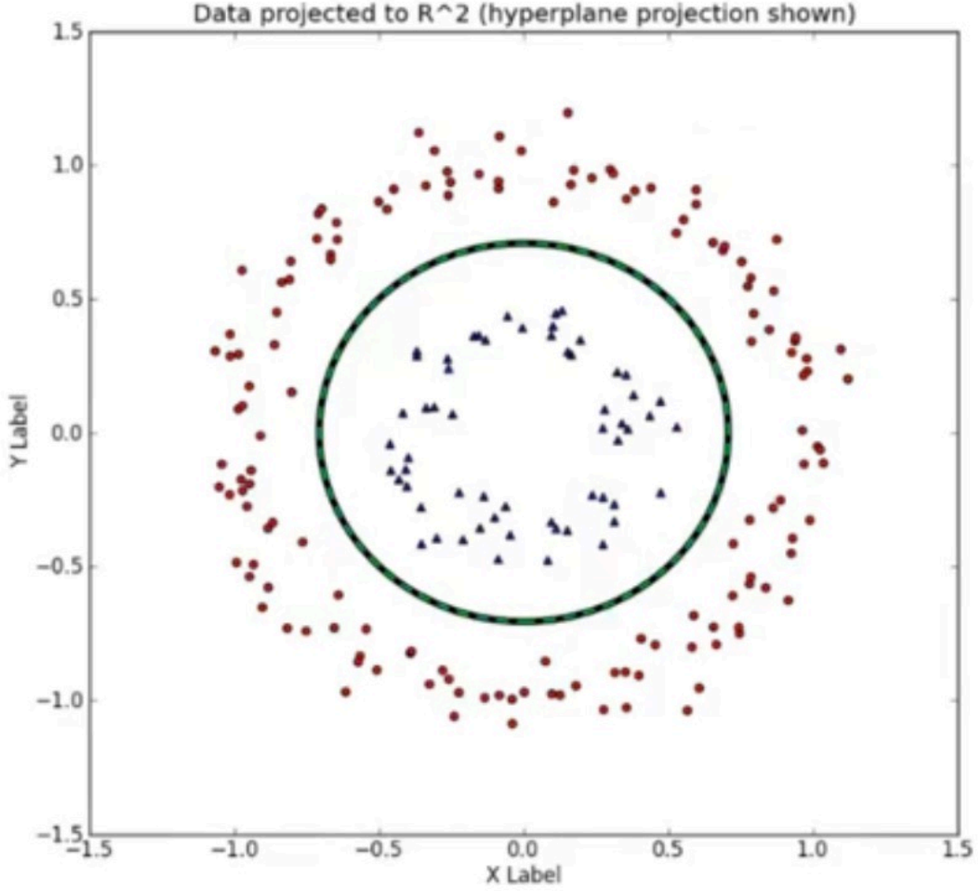
We can expand this idea to non-linearly separable data through the “kernel trick”.





Check out YouTube for nice 3D Visualization videos explaining this idea. Refer to reading for math behind this.





We'll start our example by using Support Vector Machines to predict whether a tumor is malignant or benign.

Then your portfolio project will apply these same concepts to the famous iris flower data set.

Then we'll learn how to tune our models with a GridSearch.