

Python for Data Science

Introduction to Support Vector Machines

Introduction to Support Vector Machines





Reading Assignment

Chapter 9 of
Introduction to Statistical Learning
By Gareth James, et al.



Support Vector Machines

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



Support Vector Machines

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.



Support Vector Machines

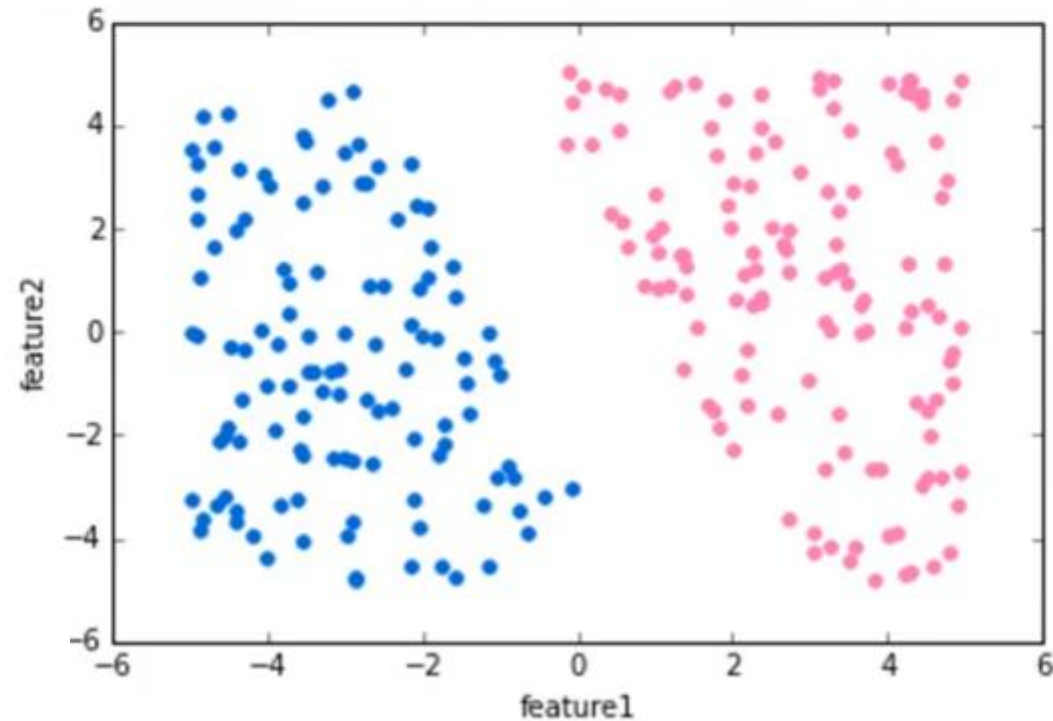
An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

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.



Support Vector Machines

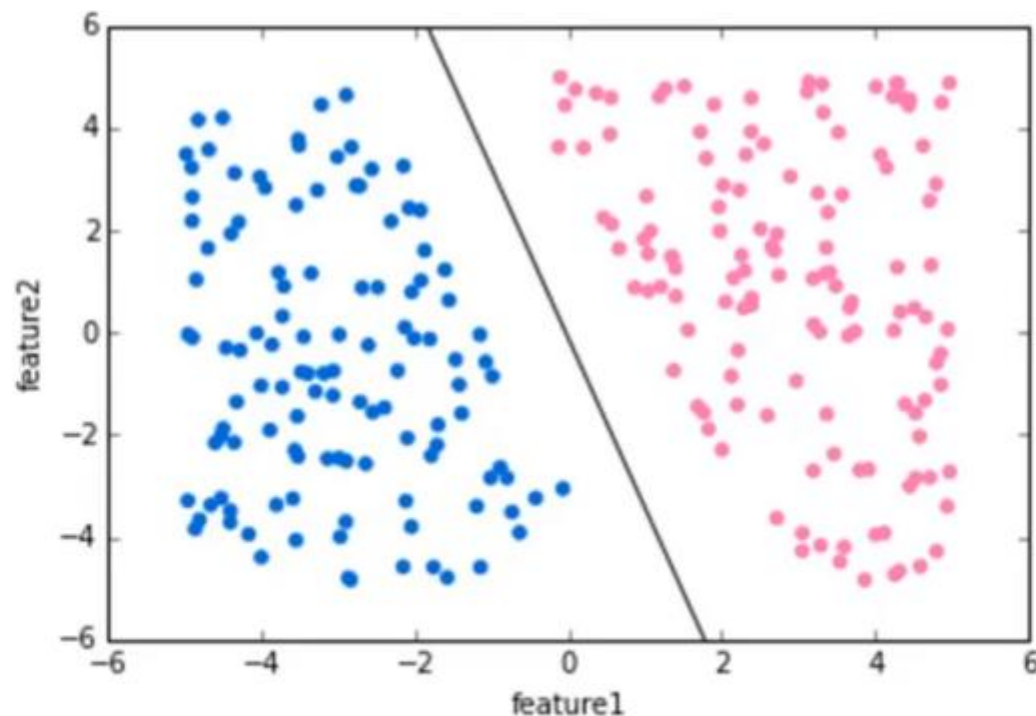
Let's show the basic intuition behind SVMs. Imagine the labeled training data below:





Support Vector Machines

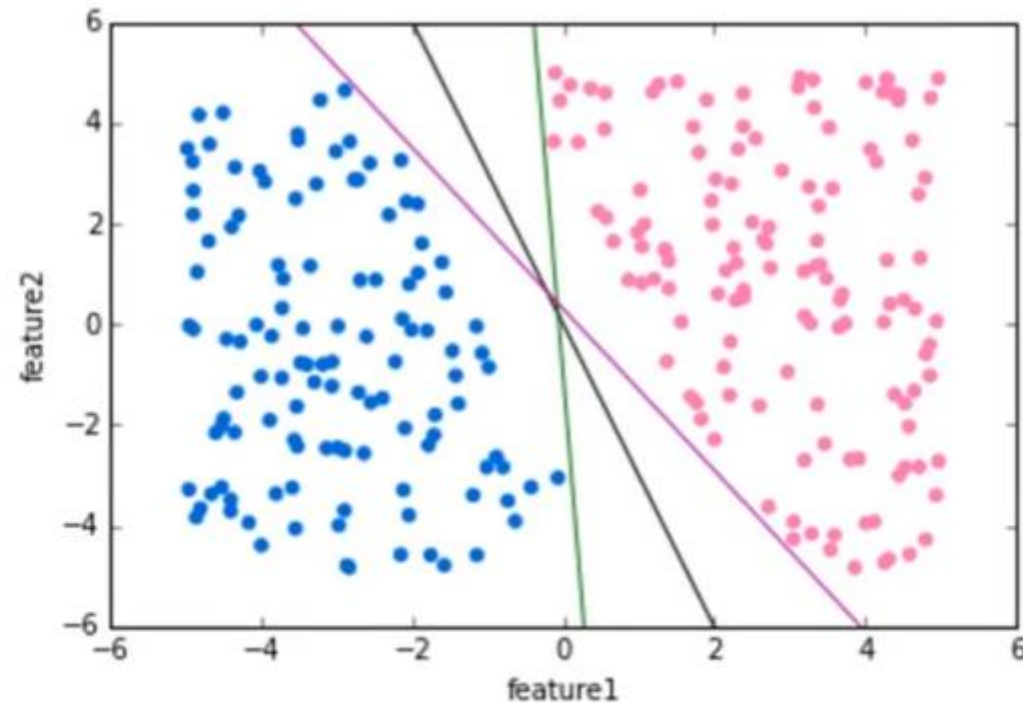
We can draw a separating “hyperplane” between the classes.





Support Vector Machines

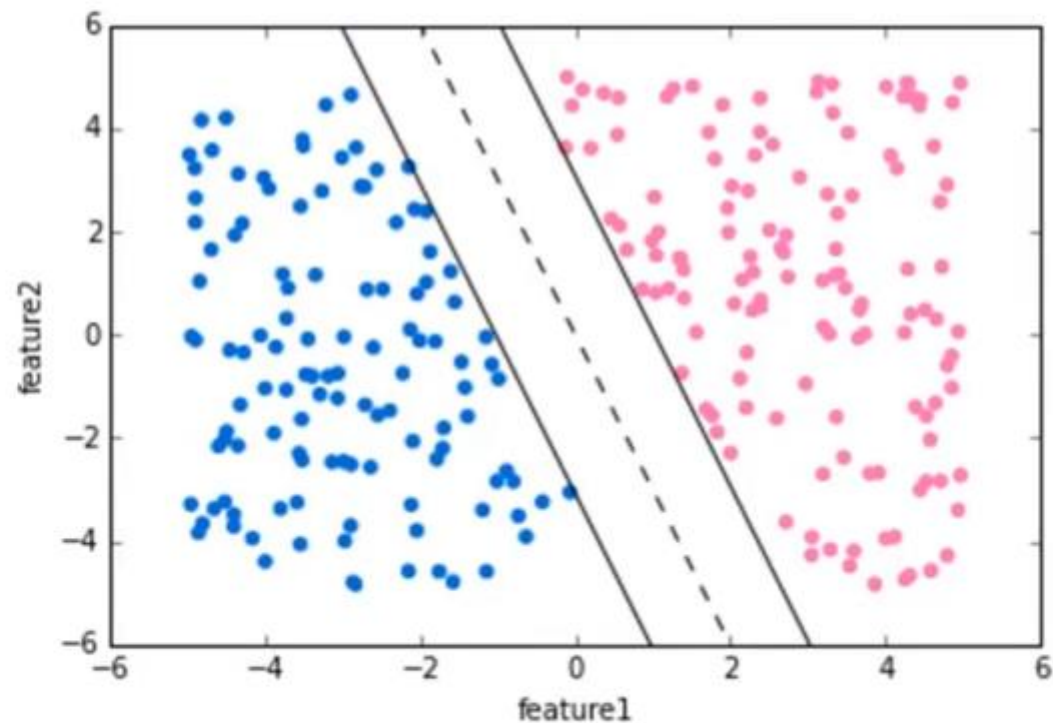
But we have many options of hyperplanes that separate perfectly...





Support Vector Machines

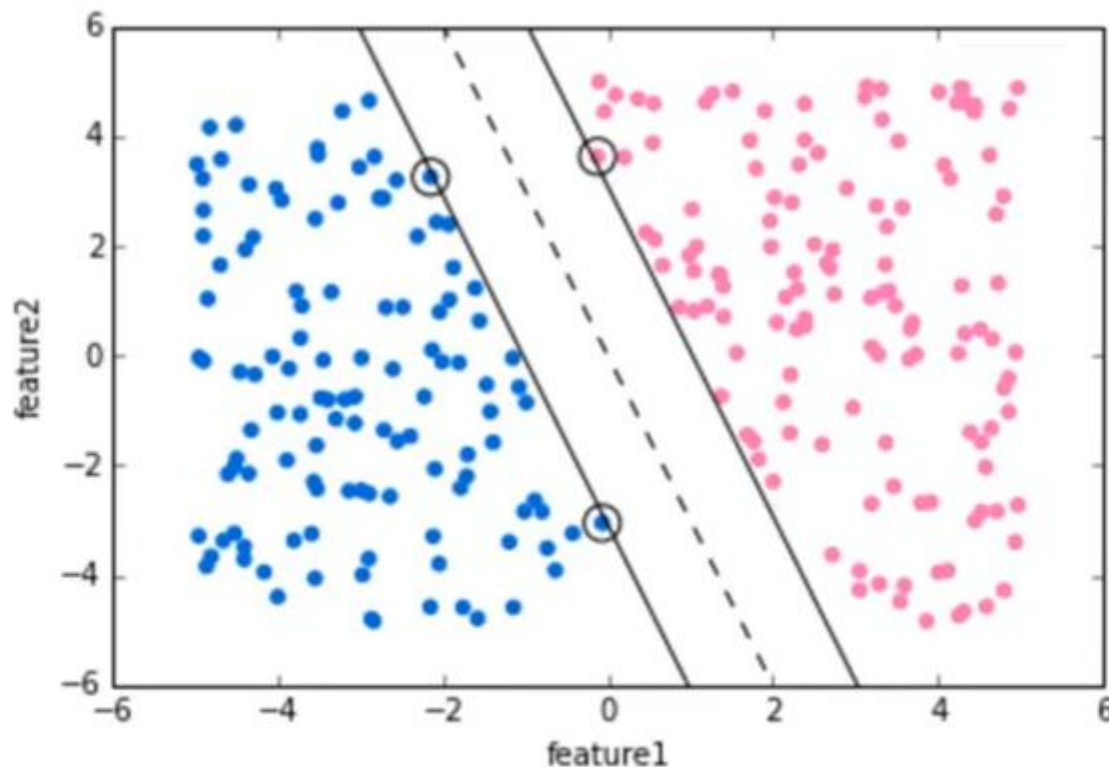
We would like to choose a hyperplane that maximizes the margin between classes





Support Vector Machines

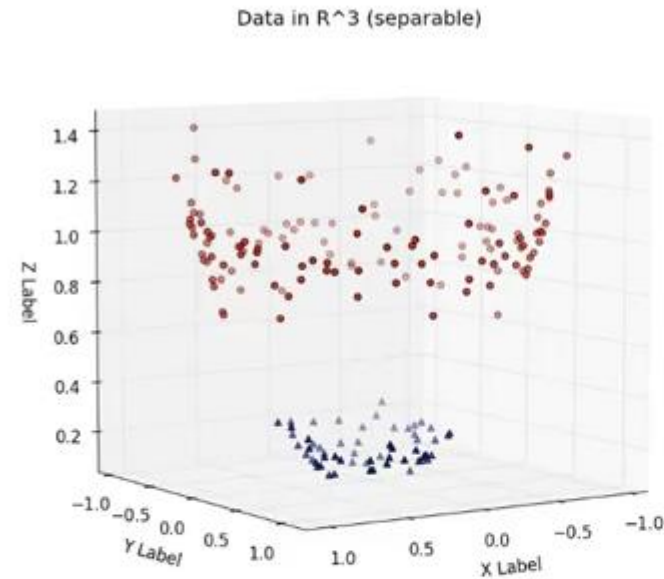
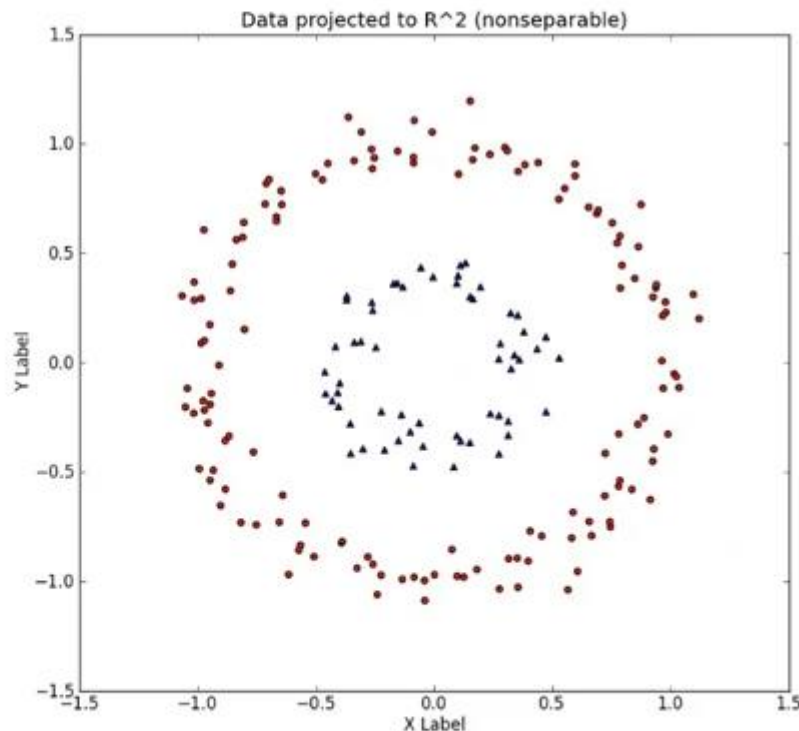
The vector points that the margin lines touch are known as Support Vectors.





Support Vector Machines

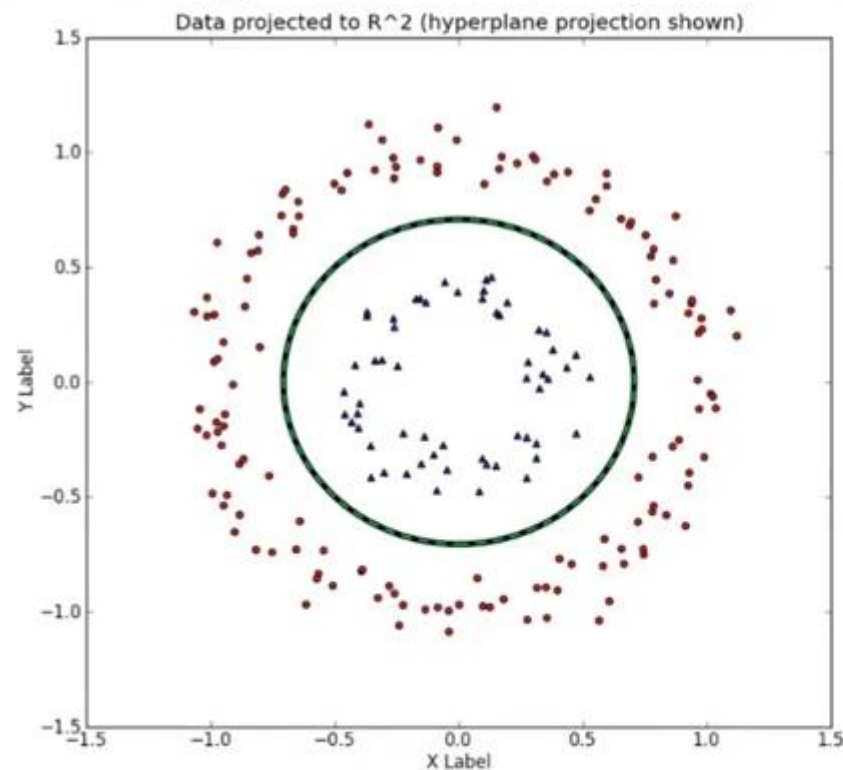
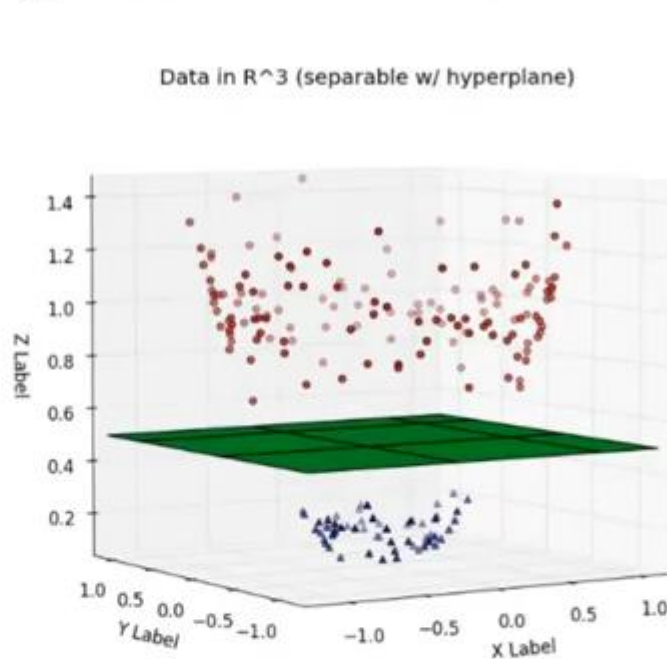
We can expand this idea to non-linearly separable data through the “kernel trick”.





Support Vector Machines

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





Example with Python

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.



Support Vector Machines With Python



Thanks!

Any questions ?