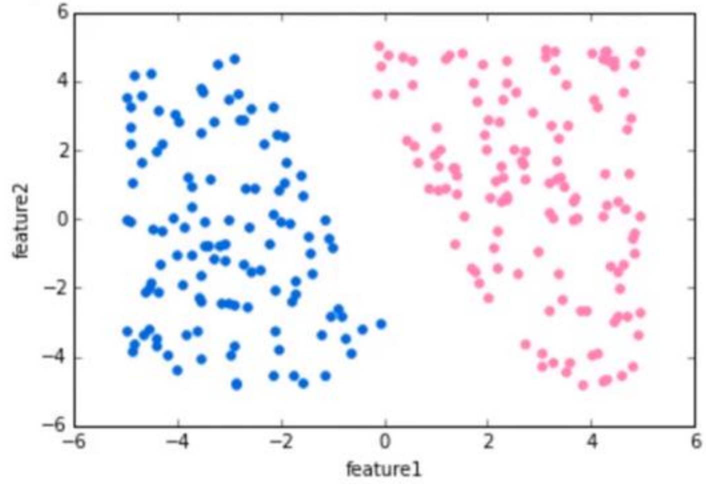
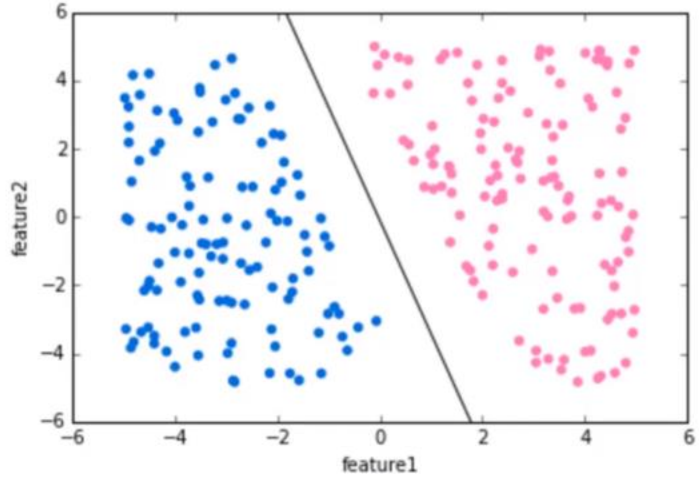
Introduction to Support Vector Machines

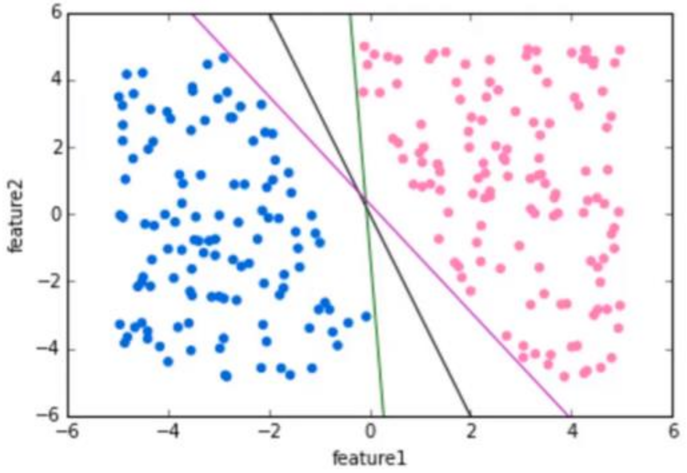
Let’s show the basic intuition behind SVMs. Imagine the labeled training data below:



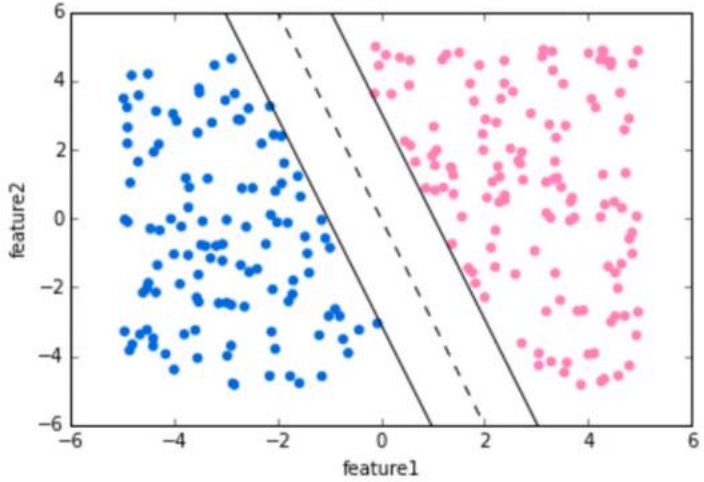
We can draw a separating "hyperplane" between the classes.



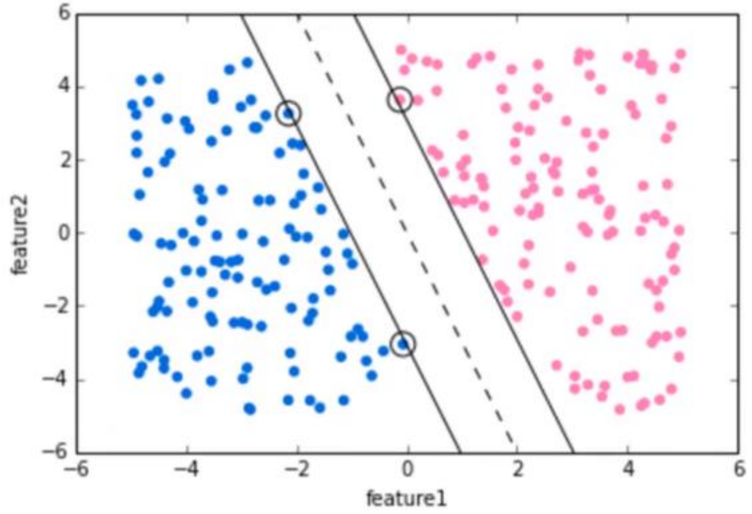
But we have many options of hyperplanes that separate perfectly...

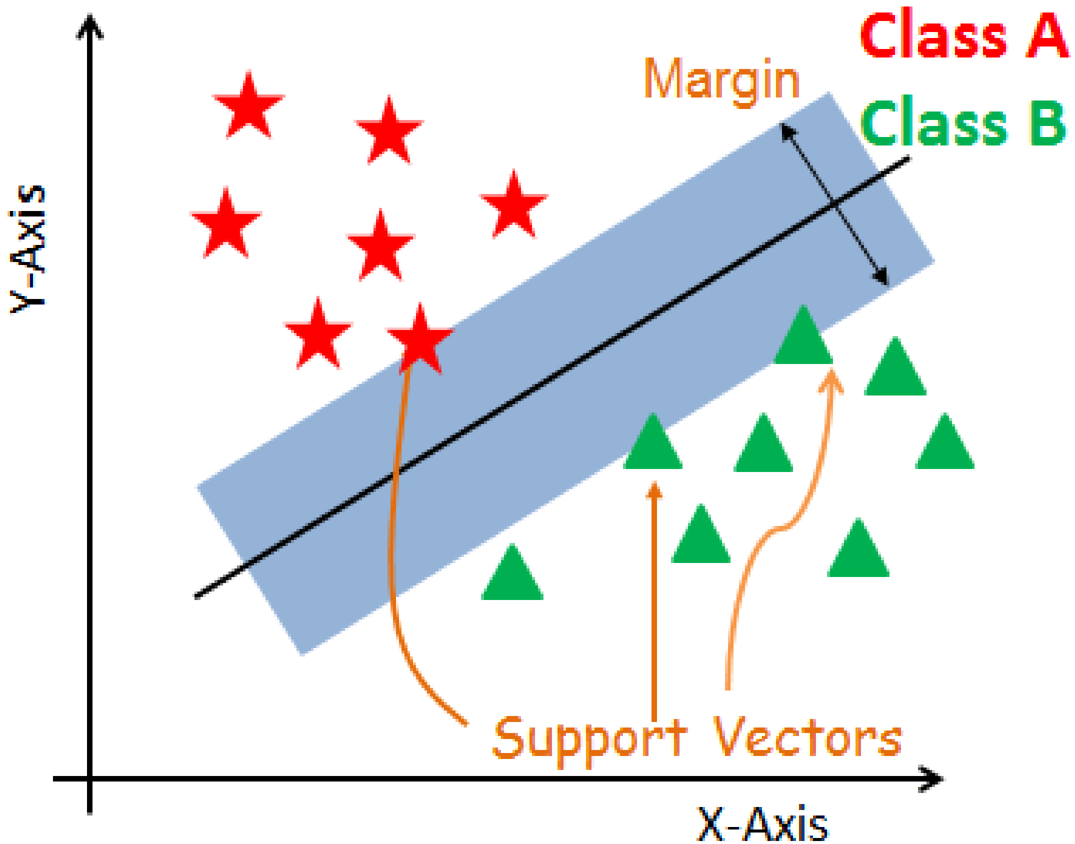


We would like to choose a hyperplane that maximizes the margin between classes

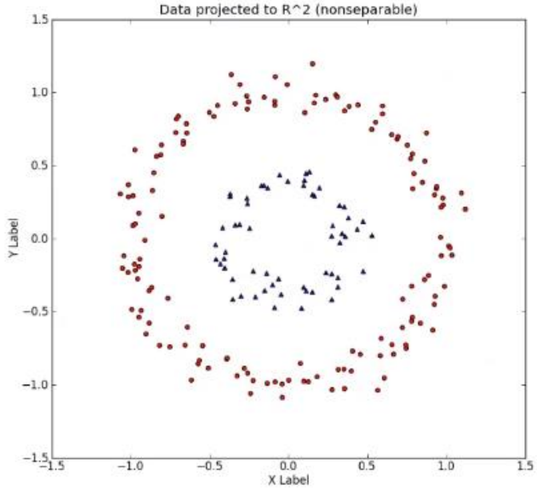


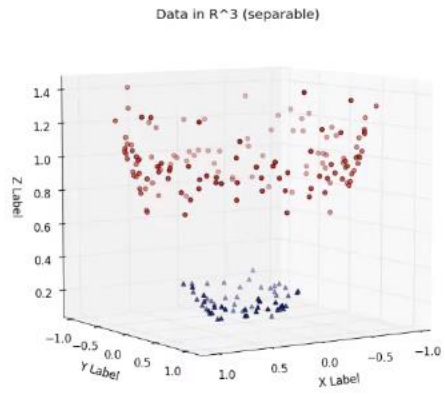
The vector points that the margin lines touch are known as Support Vectors.



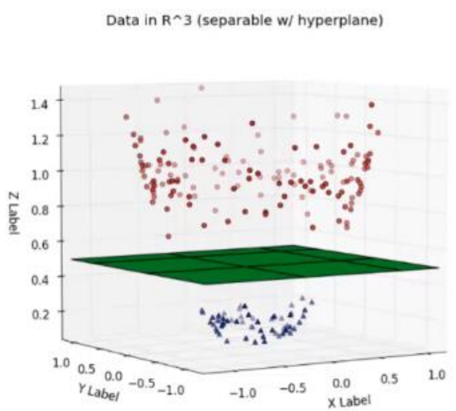

Class A
Class B
Margin
Y-Axis
X-Axis
Support Vectors

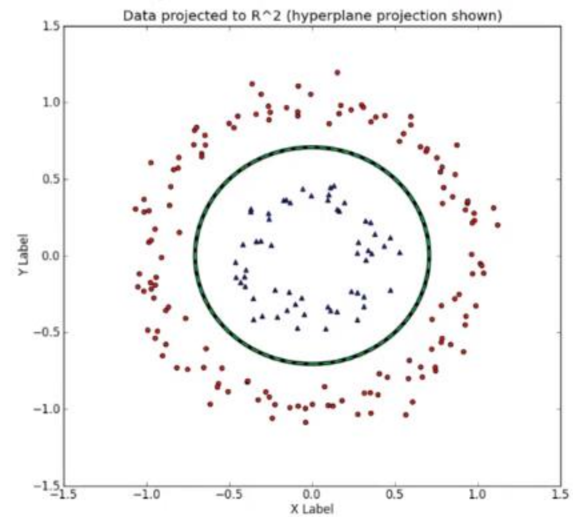
We can expand this idea to non-linearly separable data through the “kernel trick”.





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

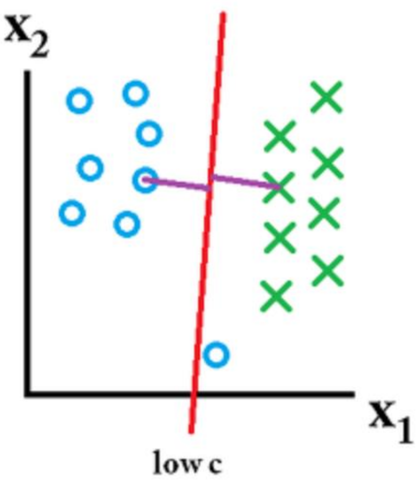


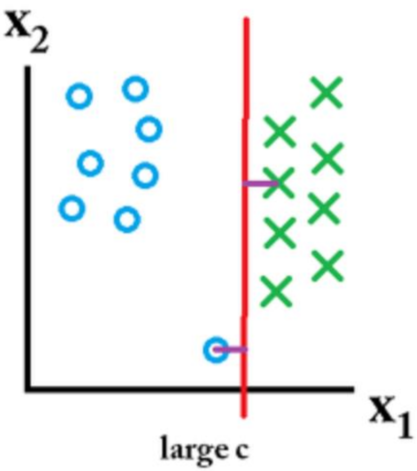


데이터를 분리하는 최적의 선형 결정경계를 찾는 알고리즘

- C

얼마나 많은 샘플이 다른 클래스로 분류되는 것을 허용할 것인가?
클수록 허용을 적게 => 과대적합 가능성
작을수록 허용을 크게 => 과소적합 가능성

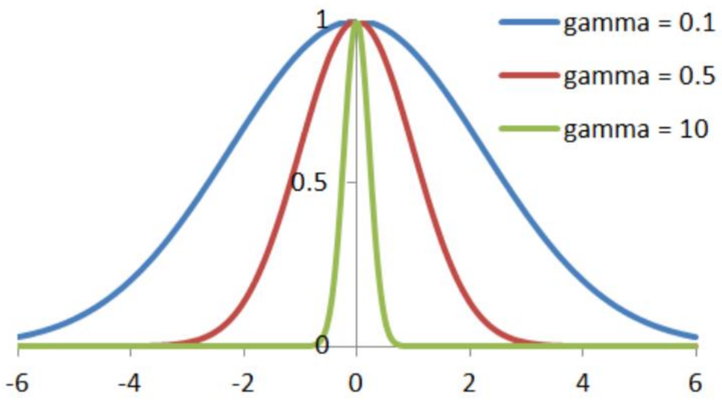


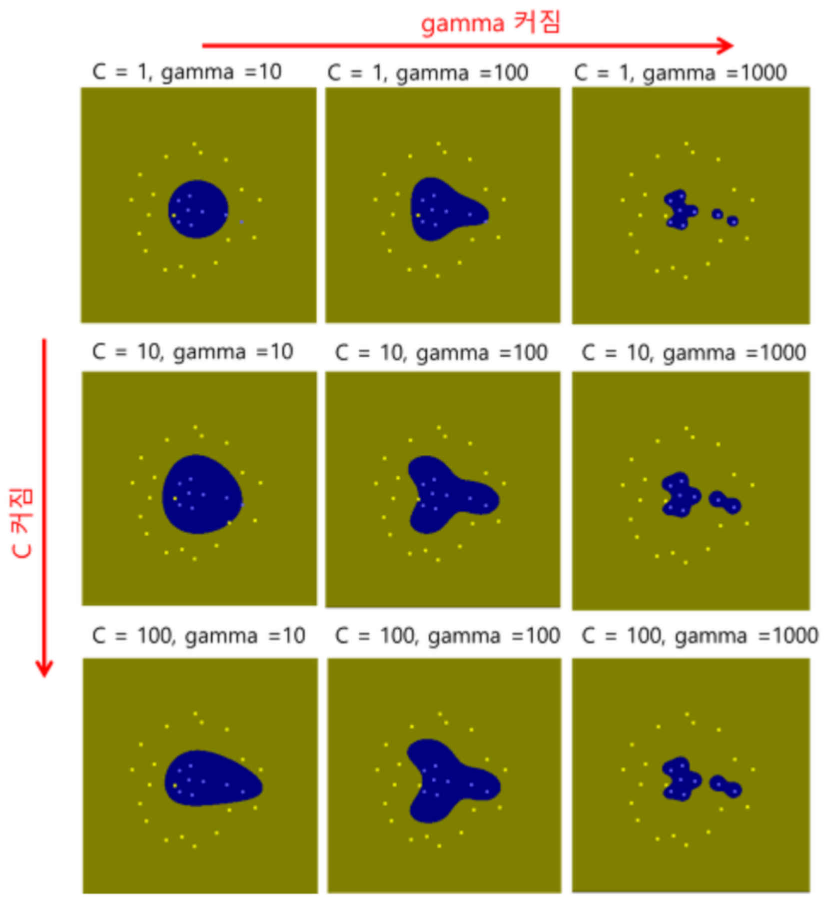


커널기법 – 주어진 데이터를 고차원의 특징 공간으로 사상한다

- Sigmoid Kernel, RBF Kernel, Ploynomial Kernel
- gamma

하나의 데이터 샘플이 영향력을 미치는 거리
클수록 영향이 좁다 => 과대적합 가능성
작을수록 영향의 범위가 크다 => 과소적합 가능성





C는 이상치의 허용범위를 결정하고
Gamma는 결정경계의 곡률을 결정한다.