

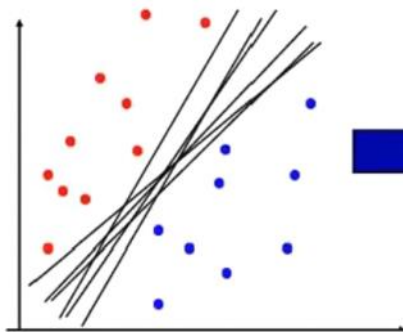
SVM(배포용)

2019년 1월 24일 목요일 오후 5:54

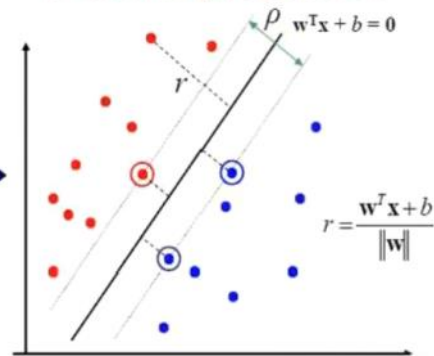
1. 개요

- Boser, Guyon 및 Vapnik에 의해 1992년 제안된 이후, 1990년대 후반부터 현재까지 학계와 업계(주로 미국 및 유럽 선진국)에서 활발하게 활용되고 있는 예측 기법
- 기계학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도학습 모델
- 주로 분류와 회귀분석을 위해 사용됨
- 두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때 새로운 데이터가 어느 카테고리에 속할지 판단하는 기준으로 가장 큰 폭을 가진 경계를 찾는 알고리즘

Which of the linear separators is optimal?



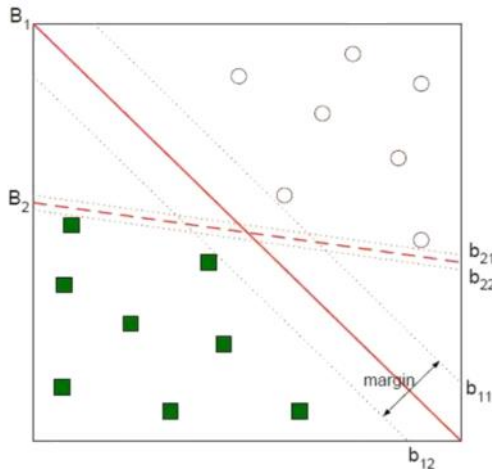
Maximum Margin Classification!



직선 B1과 B2 모두 두 범주를 잘 분류하고 있음

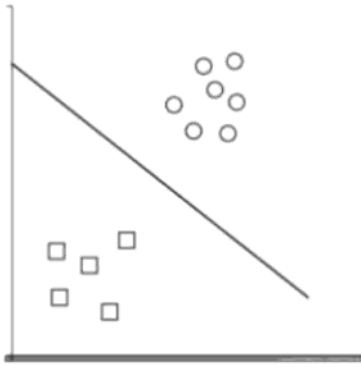
B2보다는 B1이 두 범주를 여유있게 분류하고 있으므로 좀 더 나은 분류 경계면이라고 할 수 있음

SVM은 두 범주 사이의 거리(margin)를 최대화 하는 경계면을 찾는 기법

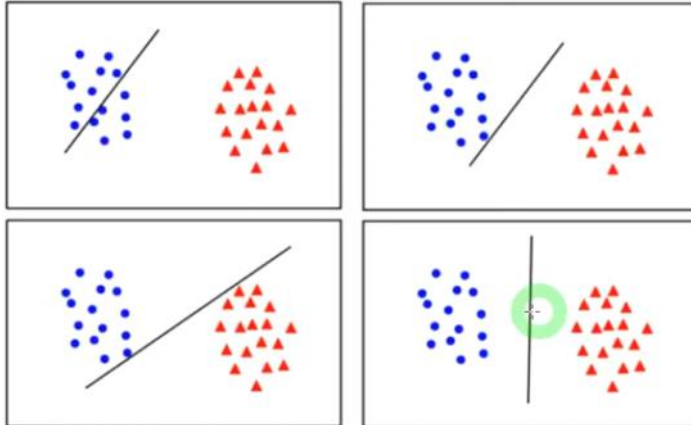


2. SVM의 분류 알고리즘

한 개체가 어느 그룹에 속하는지 결정하려면 선형 분류기를 사용하여 개체들의 위치를 알아내고 정확한 분리선(초평면)을 확인해야 한다.

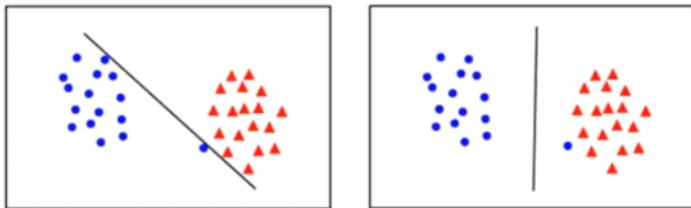


최적으로 나누는 것은 어떻게 할 수 있을까?



SVM은 margin이 가장 큰 초평면을 기준으로 분류를 수행함.

margin이 크면 클수록 학습에 사용하지 않은 새로운 데이터가 들어오더라도 잘 분류할 수 있는 가능성이 커진다.



왼쪽은 선형적으로 분류를 잘 했지만 margin이 거의 없다. 이 경우 학습에 사용되지 않은 새로운 데이터가 들어올 경우 에러가 발생할 가능성이 높다.

오른쪽은 파란색 점 1개에 대해 오분류를 하기는 했지만 margin은 매우 크다.

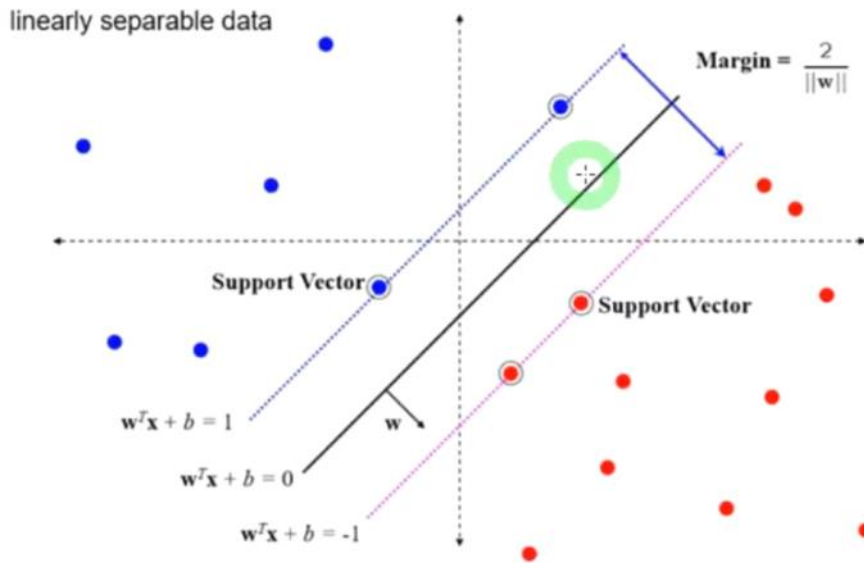
그 점이 특이점이라면 차라리 오분류하는 것이 나을 수도 있음.

margin이 충분하기에 새로운 데이터에 대한 예측은 더 잘 할 수도 있다.

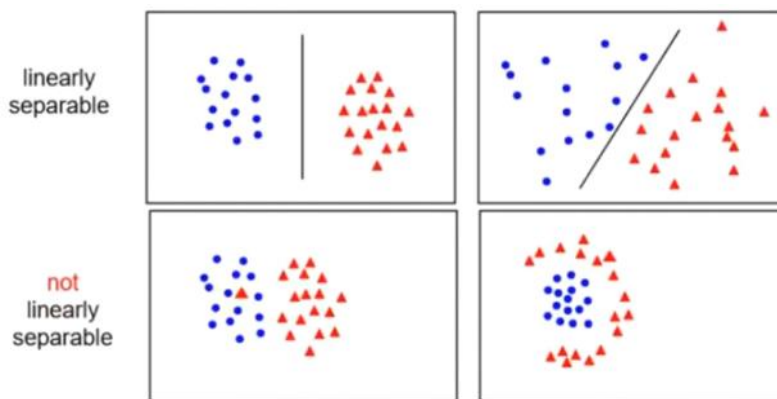
파란색 점과 빨간색 점을 최대 마진으로 구별하는 직선은 검은색 직선이다.

여기서는 2차원이기 때문에 직선이지만 3차원 이상이 되면 평면이 되며 이것을 초평면(hyper-plane)이라고 한다.

또한 hyper-plane으로부터 가장 가까이 있는 파란색 점과 빨간색 점을 Support Vector라고 한다.



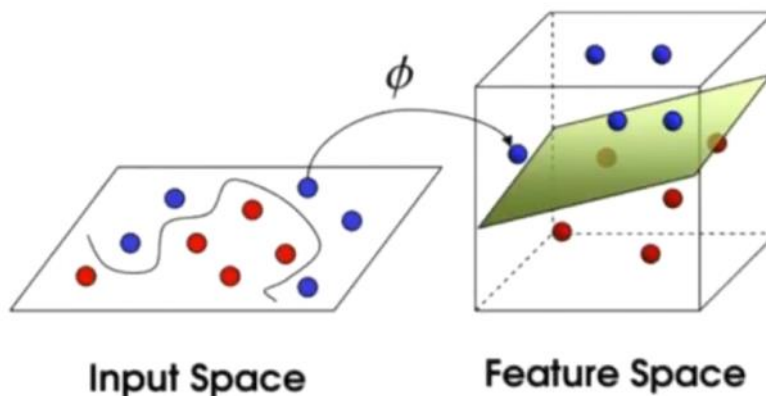
선형적으로 분류가 가능한 경우와 그렇지 못한 경우



아래 그림과 같이 선형적으로는 도저히 구별하기 어려운 특징들에 대해서도 초평면을 구할 수 있을까?



아래 그림처럼 구별이 가능한 방향으로 mapping을 시키면 원래는 선형적으로 구별이 불가능한 특징들도 새로운 공간에서는 구별이 가능하게 된다.



위와 같은 비선형 분류 문제를 해결하기 위해 SVM에서는 커널 함수(또는 커널 트릭이라고도 함)를 사용한다.

3. SVM의 장점

- 선형 분류 뿐 아니라 비선형 분류도 가능
- 모델을 만들 때 고려해야 할 파라미터가 많지 않음

- c. 적은 양의 데이터로도 모델을 만들 수 있음
- d. 딥러닝이 이전에는 분류 모형 중에서 기술적으로 가장 진보된 모형으로 평가 되었음

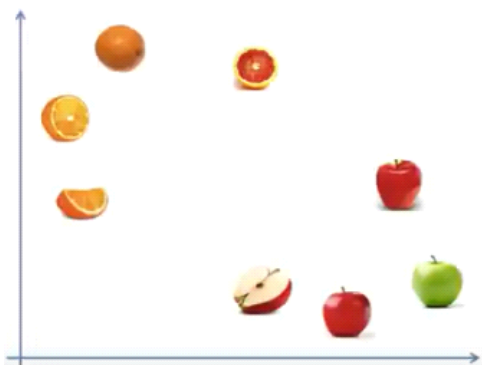
4. SVM의 적용분야

- a. 문서 분류
- b. 이미지 인식, 안면 인식
- c. 주가 예측
- d. 부도 예측
- e. 고객 구매 예측
- f. 채권 등급 평가 예측
- g. 필체 패턴 인식

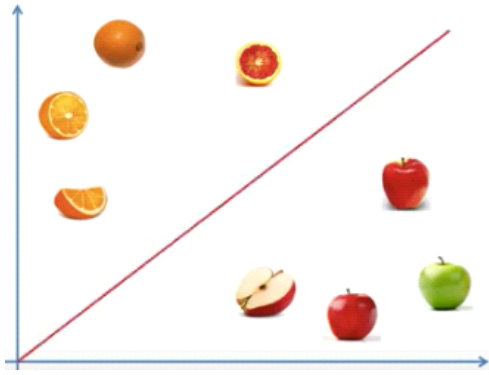
What is SVM?

Support Vector Machine

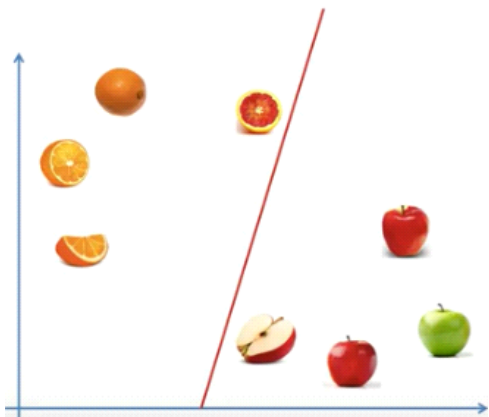
apple, orange classifier



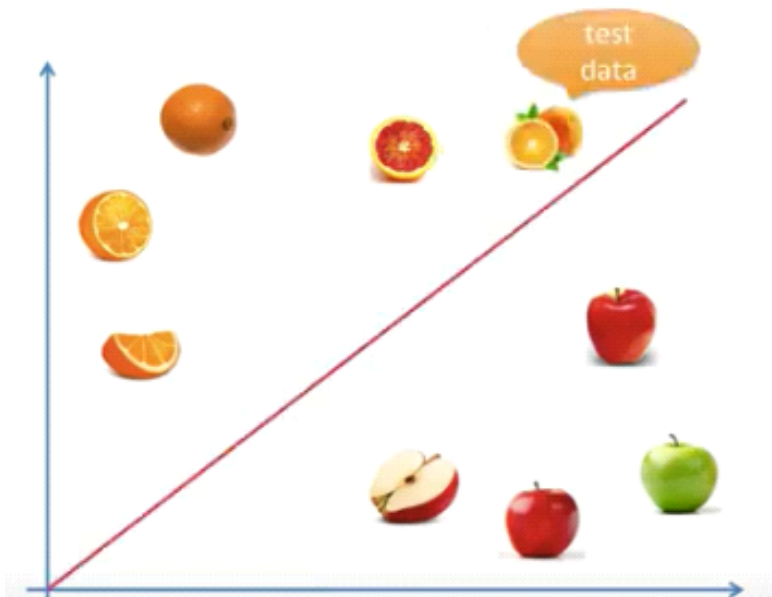
apple, orange classifier



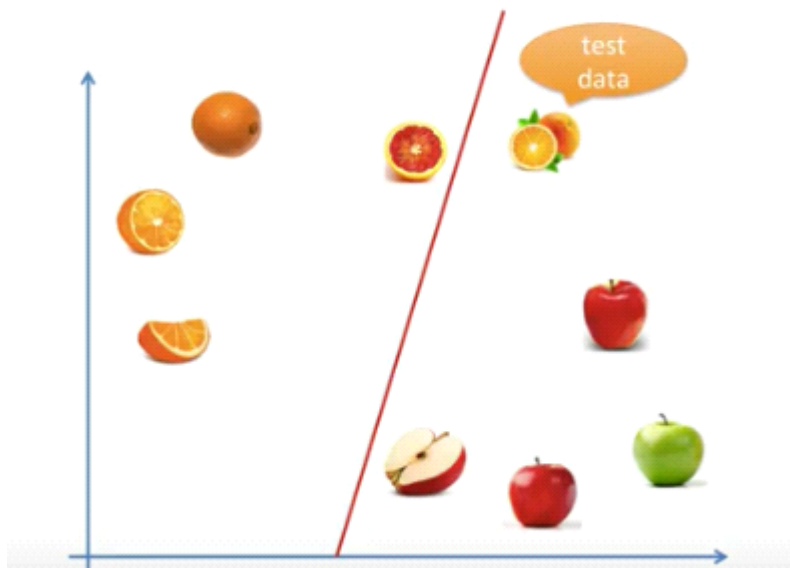
apple, orange classifier



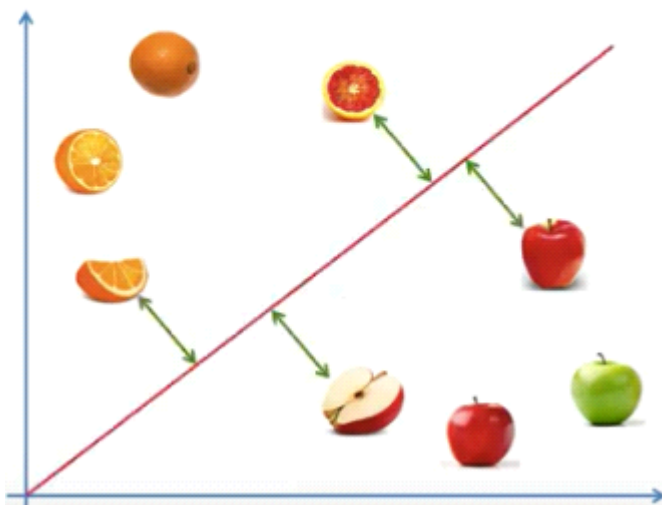
test data predicted well!



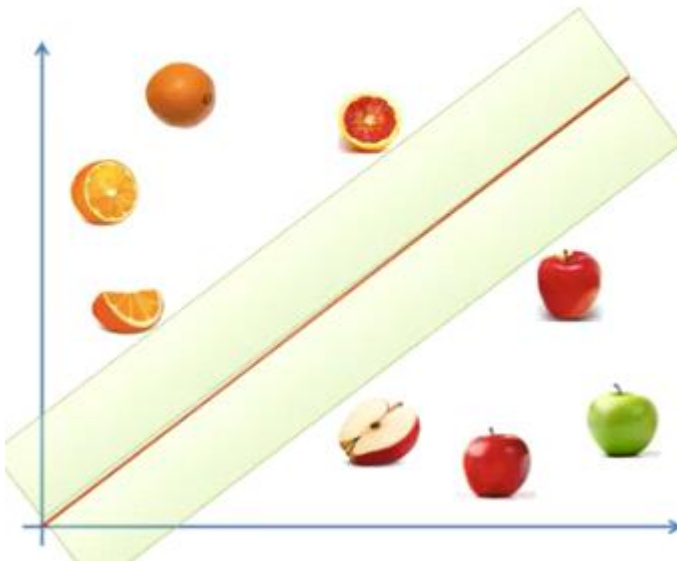
test data prediction wrong!



margin



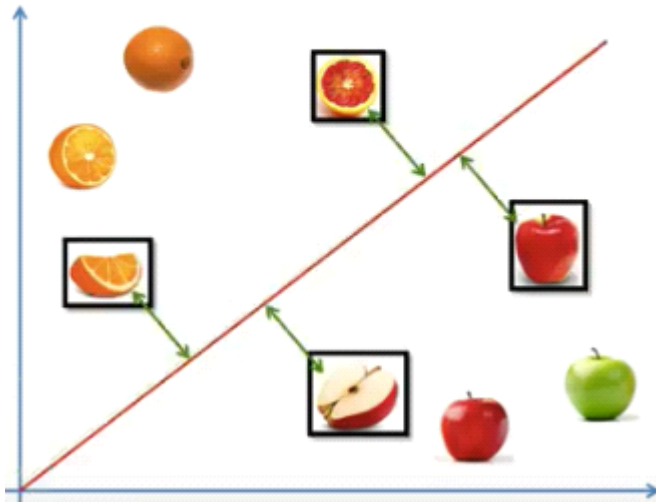
margin



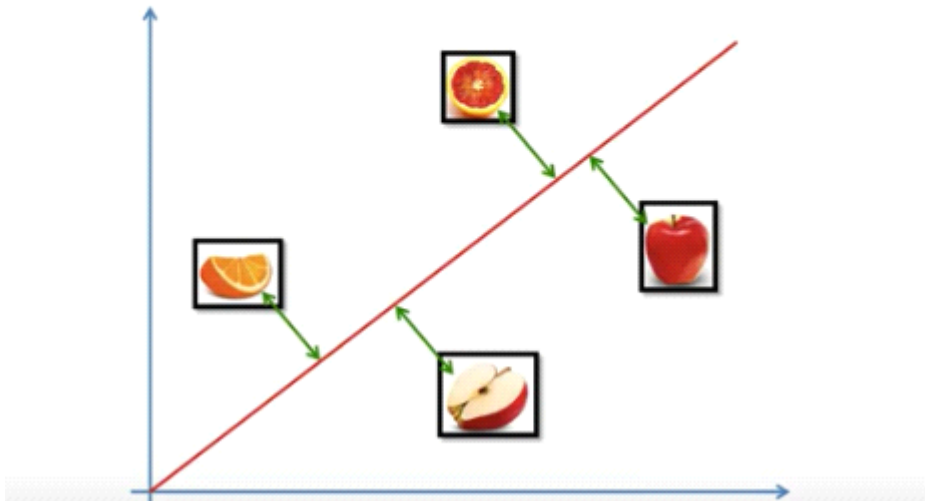
margin



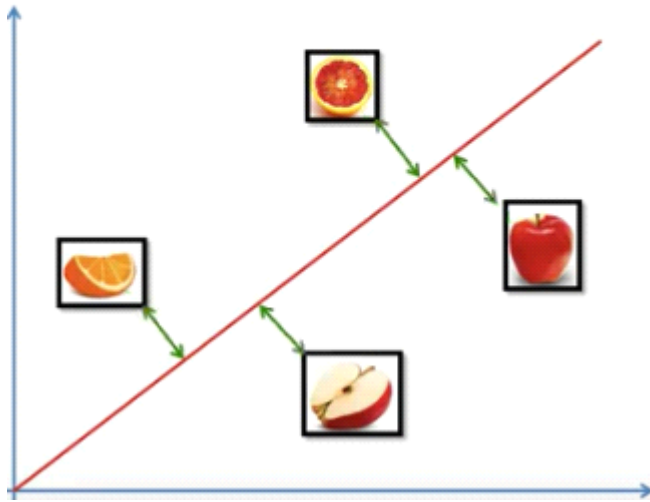
support vector



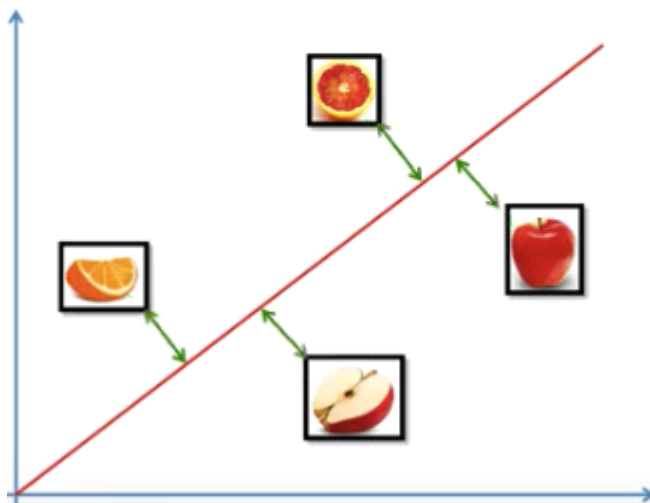
SVM only uses support vector for prediction



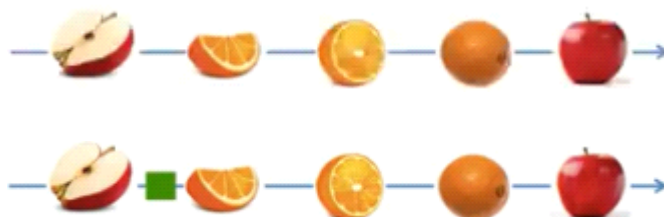
less computation!

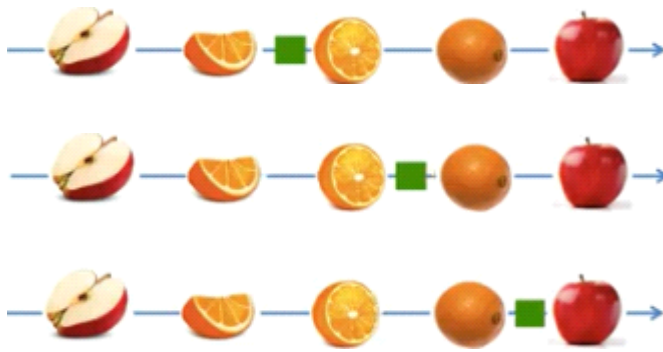


This is linear SVM



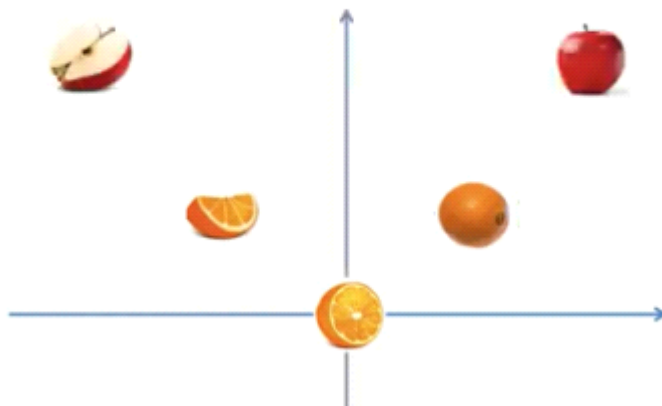
What if data is not linearly separable?



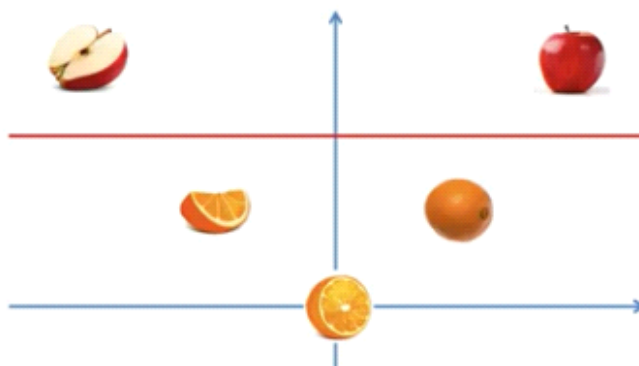


mapping lower dimension to high dimension

$$y = x^2$$



Now it is linearly separable in higher dimension



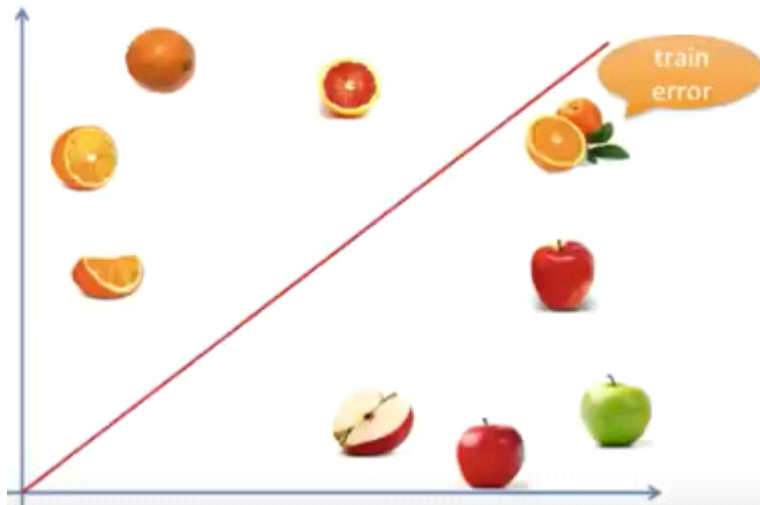
mapping to high dimension requires much computation!

Kernel trick in SVM do this without explicitly move data point to higher dimension!

]

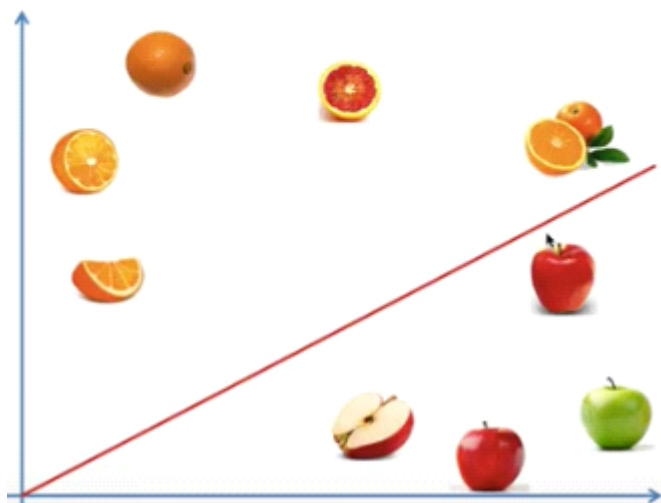
SVM Parameter1 – Cost

when cost is small

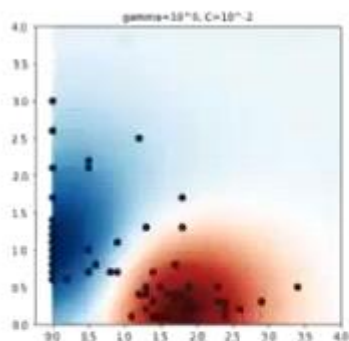


SVM Parameter1 – Cost

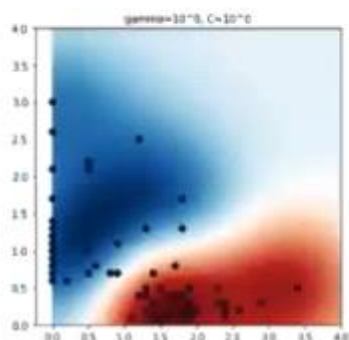
when cost is high



SVM Parameter1 – Cost

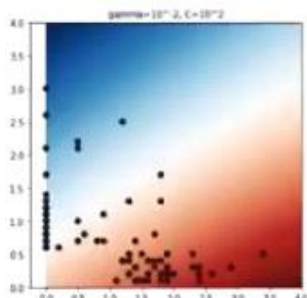


cost= 0.01

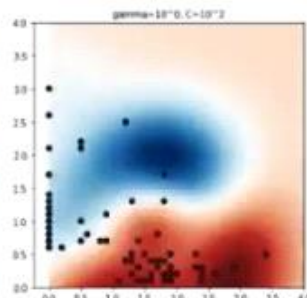


cost= 1

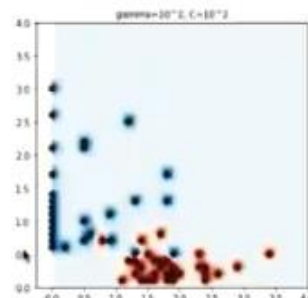
SVM Parameter2 – Gamma



gamma = 0.01



gamma = 1



gamma = 100

Grid Search – find optimal parameters

		Gamma		
		1	10	100
cost	1	0.7	0.8	0.7
	10	0.8	0.8	0.9
	100	0.6	0.8	0.8

