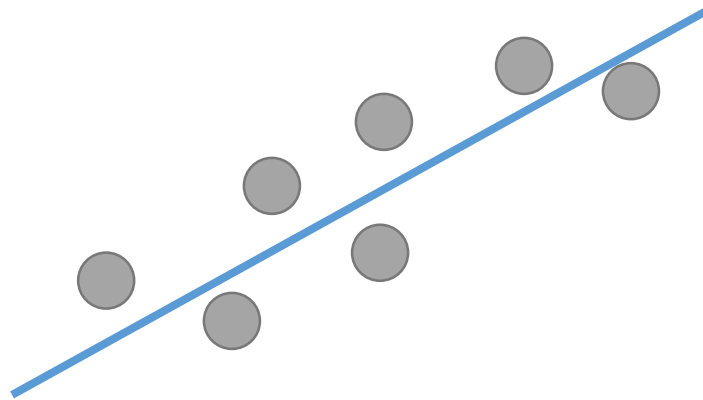


Support Vector Machine

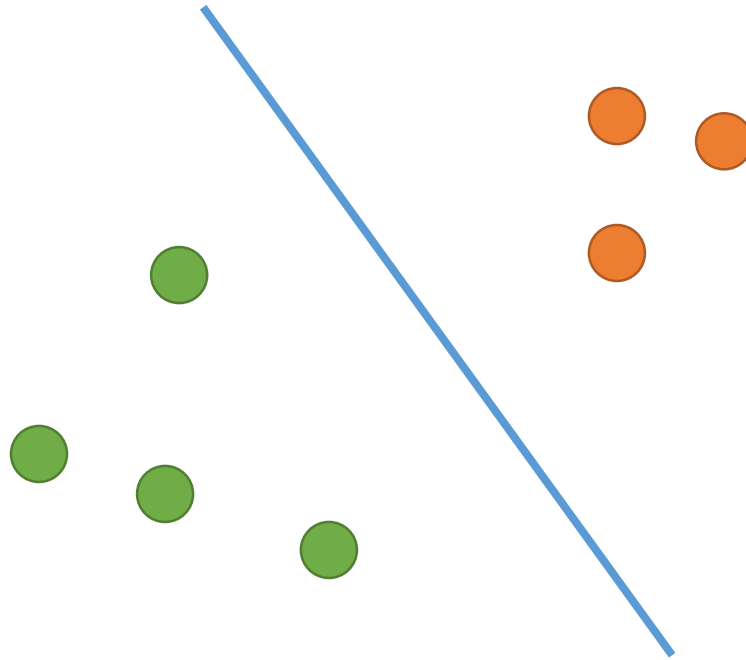
회귀

- 연속 변수를 예측
- 데이터를 지나는 추세선을 찾는 것



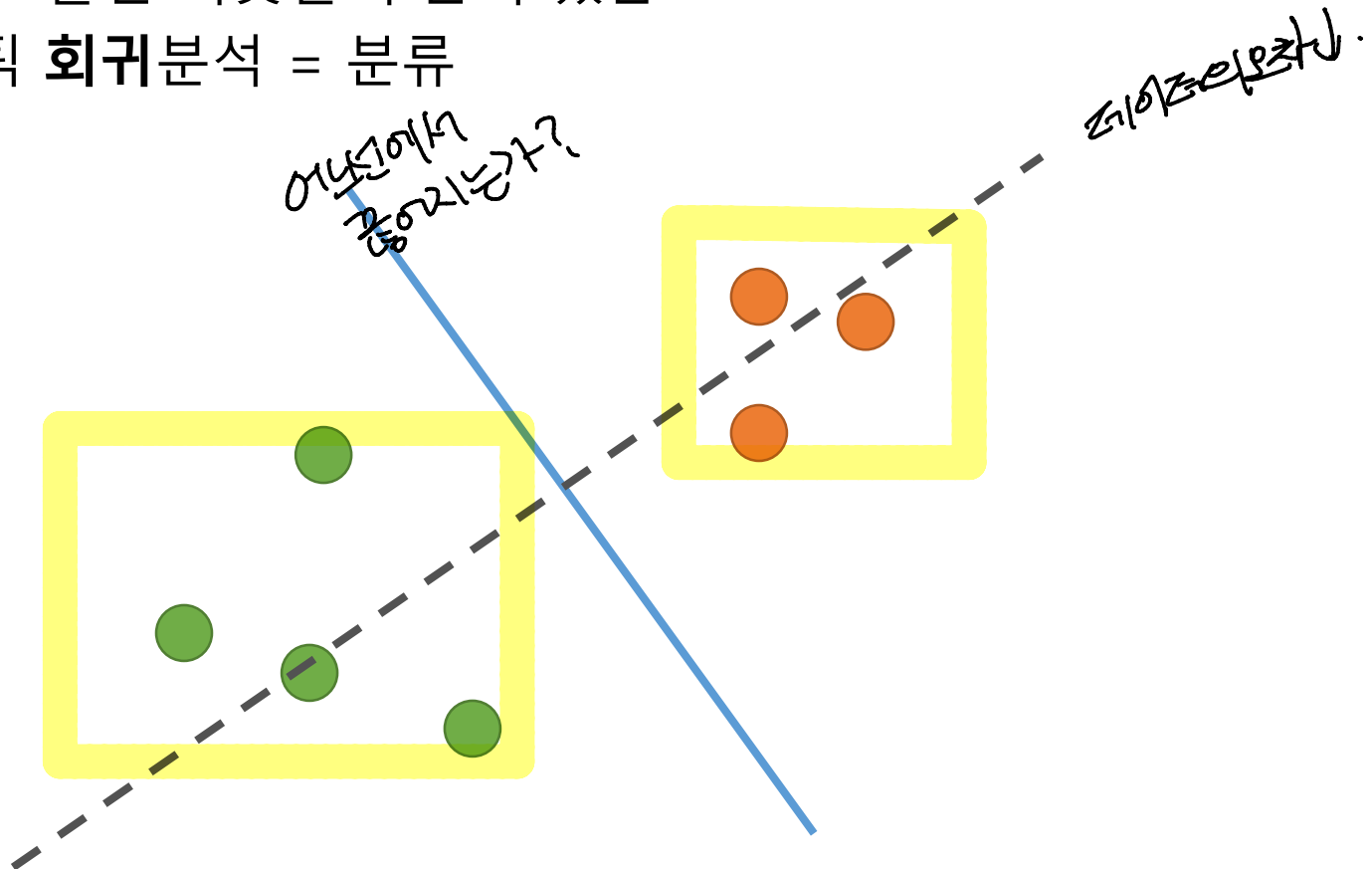
분류

- 이산 변수 또는 범주형 변수를 예측
- 데이터를 나누는 **경계선**을 찾는 것



회귀와 분류는 다른 것인가?

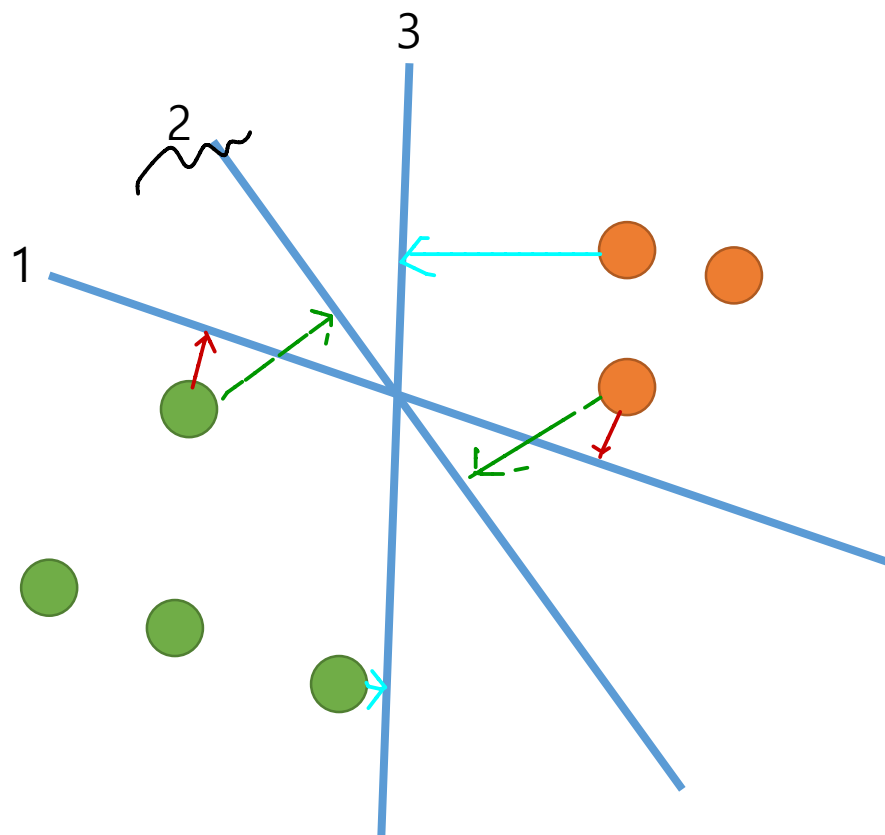
- 깊이 파고들면 비슷한 부분이 있음
- 로지스틱 **회귀**분석 = 분류



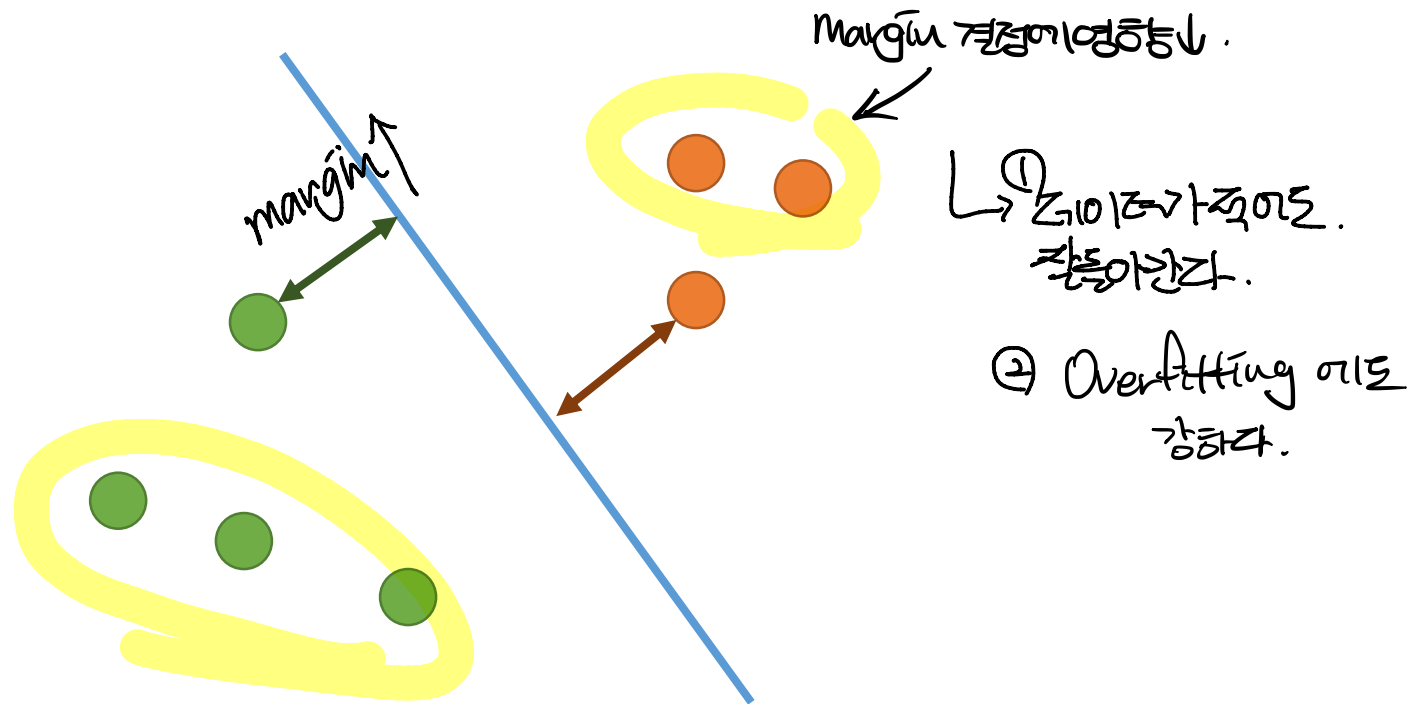
Support Vector Machine

- 한동안 인기를 구가하던 기계학습 모형
- 선형 모형: 오차를 줄이는 데 관심
- SVM: 좋은 형태를 찾는데 관심

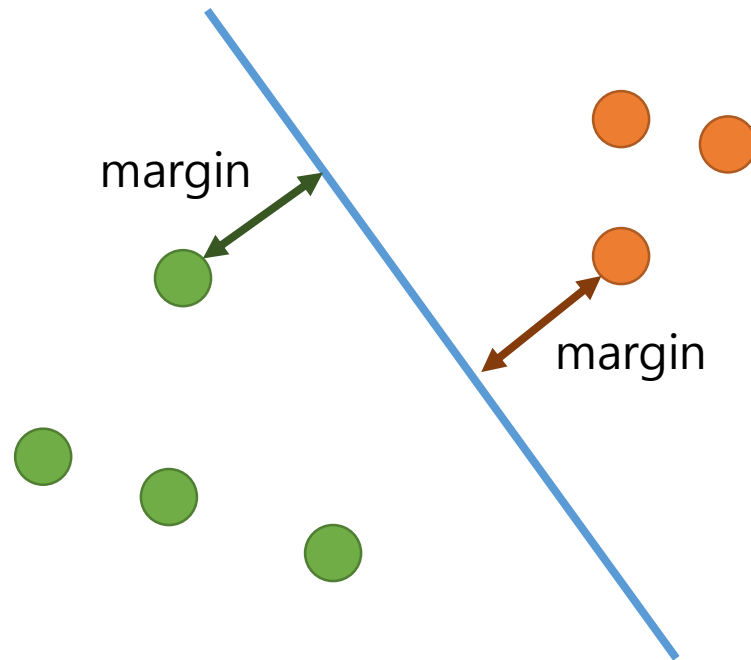
무엇이 좋은 형태인가?



왜 2번인가?

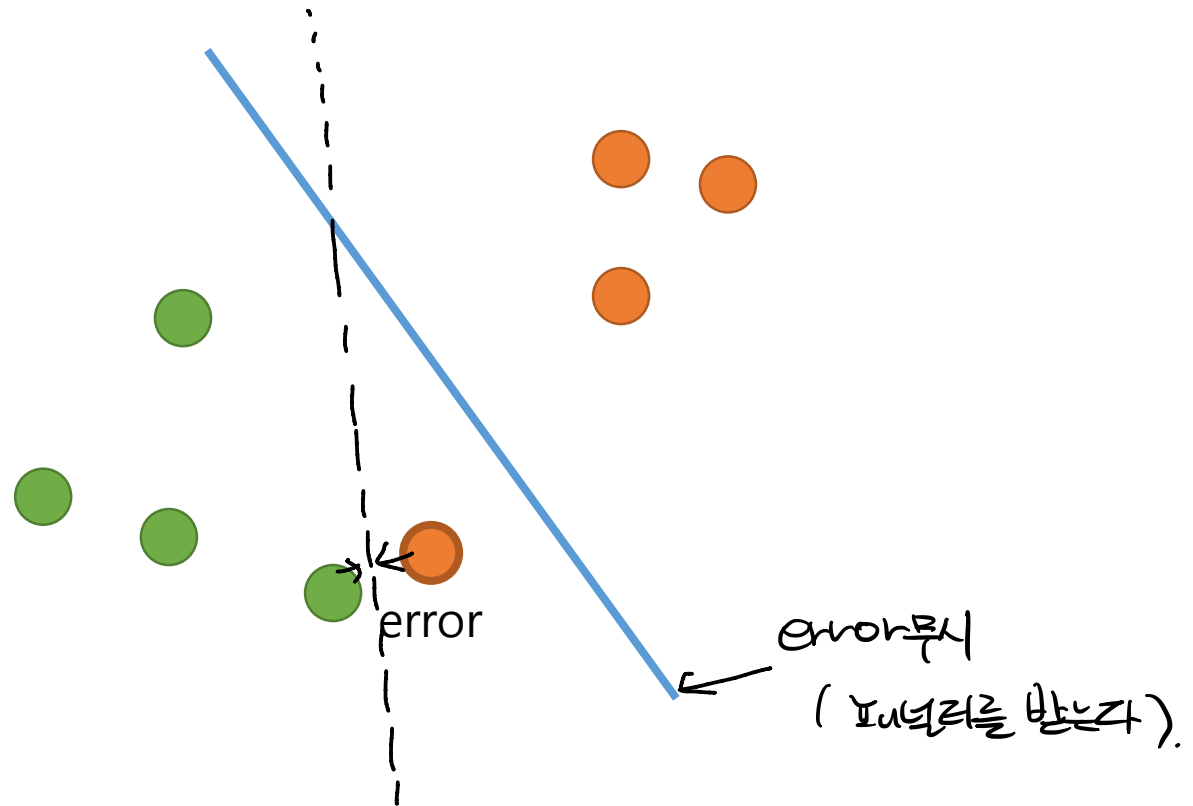


Large Margin Classifier



이런 경우는?

- 오차에는 페널티를 부과한다

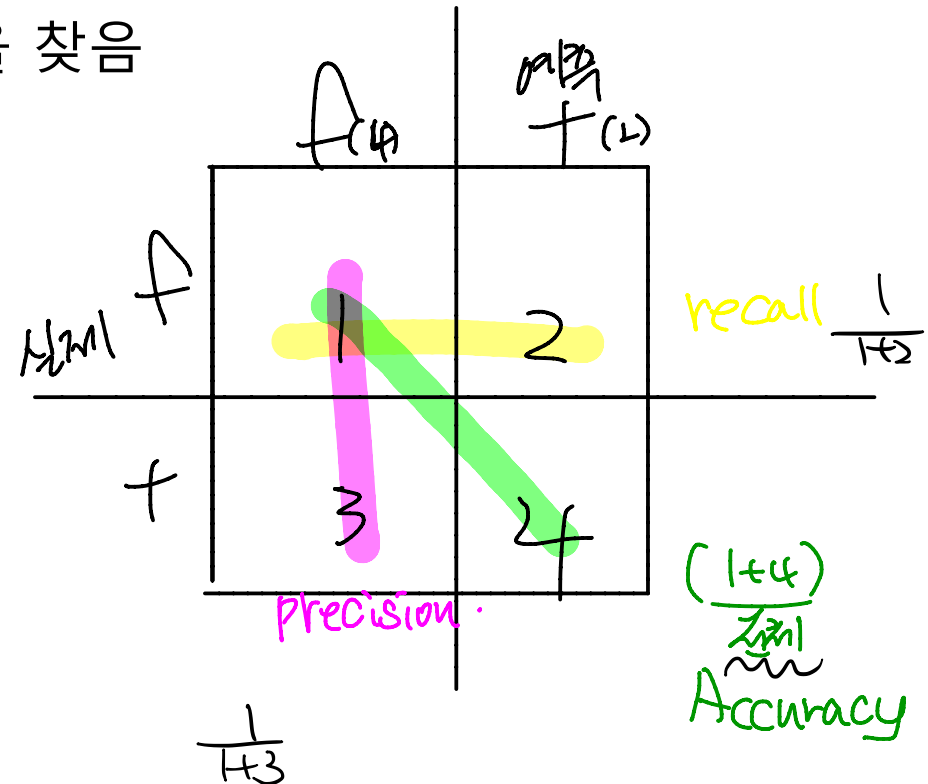


SVM과 선형 모형의 관계

- 정규화 선형 모형 \rightarrow 오차를 줄이자 + 좋은 형태도 찾자
- SVM \rightarrow 좋은 형태를 찾자 + 오차도 줄이자
- SVM은 Ridge 선형 모형과 수학적으로 비슷

Support Vector Regression

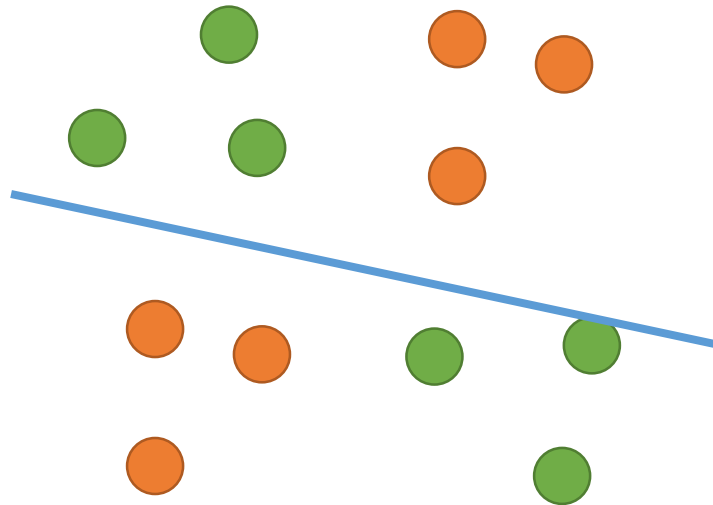
- SVM의 회귀 버전
- 가능한 평평한 형태의 추세선을 찾음



커널 트릭(kernel trick)

비선형 문제

- 현실의 많은 문제들은 비선형성(non-linearity)이 있음
- 아래 두 집단의 선형 경계선을 찾을 수 없음



XOR 문제

- XOR: 자연어에서 '또는'과 비슷. A거나 B, 둘 다는 안됨
- 커피 xor 콜라 → 둘 중에 하나만 마셔야 함
- XOR은 비선형 문제

비선형 모형

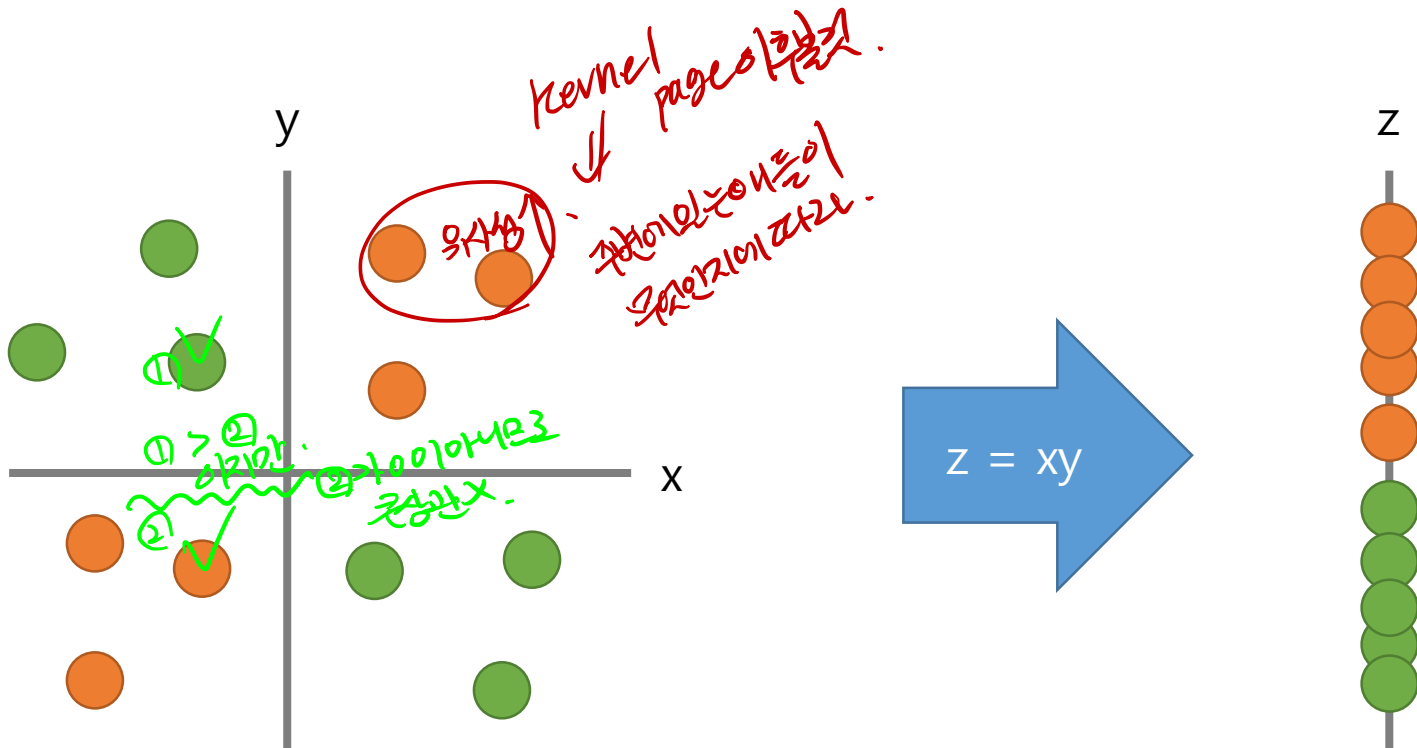
- 비선형 문제는 비선형 모형으로 풀 수 있다
- 그러나 비선형 모형은 적합(fitting) 시키기가 매우 어려움

한 가지 아이디어

- 모델을 비선형으로 만들 수 없다면

ML (Andrew) \Rightarrow feature change.

- 데이터를 비선형으로 변환 \rightarrow 선형 분리 가능하게 만들자



커널 트릭

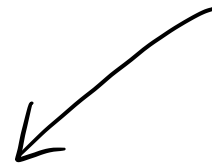
- 비선형 변환도 쉽지 않음

- 마치 비선형 변환을 한 것처럼 하자

- 유유상종: 비슷한 것들은 가까이 있다

- 가까움(거리)을 재정의 → 마치 비선형 변환을 한 것처럼 작동

SVM 거리를 재는 것이 함수만에
존재한다. \Rightarrow 거리를 재정의



커널의 종류

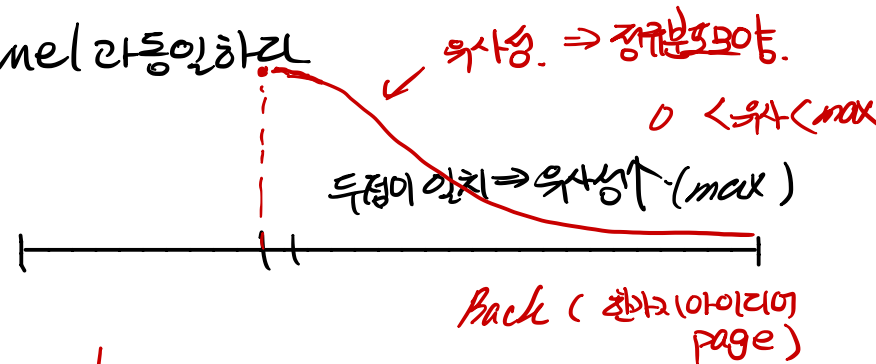
• 선형 커널(linear kernel) \Rightarrow No kernel 라 동일하리

• RBF 커널(radial basis function)

• 다항 커널(polynomial kernel)

• 시그모이드 커널(sigmoid kernel)

• 커널을 결합해도 커널이 됨



\hookrightarrow 거리에 따르는 우사성을 어떻게 조절하겠인가?

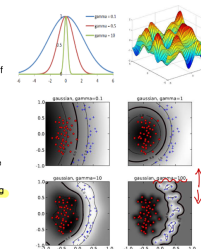
\hookrightarrow gamma. \downarrow 탄산 \Rightarrow SD \uparrow 넓은.

\uparrow 압박 \Rightarrow SD 좁은

SVM - considered heuristics

Gamma parameter (γ)

- The larger gamma the thinner gaussian curve is.
- Gaussian kernelling for a point x consists in calculation of a sum of Gaussian "bumps" centered around each support vector in a training dataset.
- For small values of γ (upper left) the decision boundary is nearly linear.
- As γ increases the flexibility of the decision boundary increases.
- Large values γ of lead to overfitting (bottom).



$$\gamma = \frac{1}{SD}$$

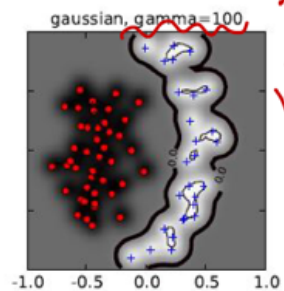
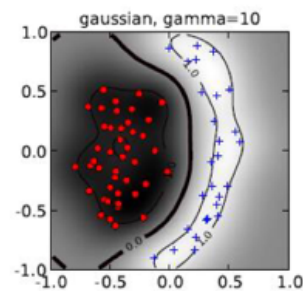
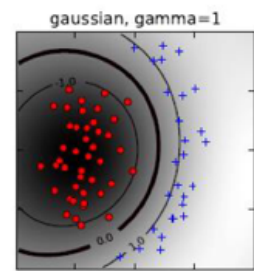
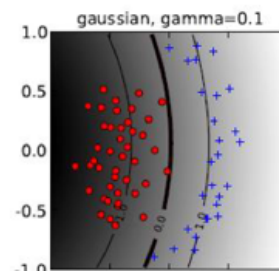
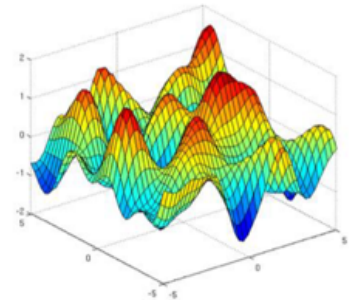
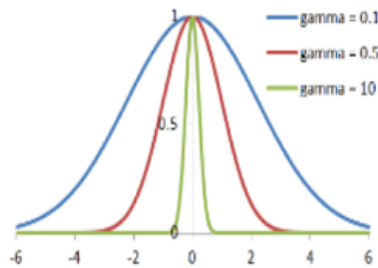
정분포 모양

SD \uparrow 탄산.
 \Rightarrow gamma \downarrow

SVM – considered heuristics

■ Gamma parameter (γ)

- The larger gamma the thinner gaussian curve is.
- Gaussian kerneling for a point x consists in calculation of a sum of Gaussian "bumps" centered around each support vector in a training dataset.
- For small values of γ (upper left) the decision boundary is nearly linear.
- As γ increases the flexibility of the decision boundary increases.
- Large values γ of lead to overfitting (bottom).



↑ ⇒ 과적합
↓ ⇒ rough

커널 트릭의 장점

저장 공간이.

ex) 답안을 수리로?

↳ 유사성만 계산할 수 있다면?

⇒ 같은 글자 몇 개인가?

거리만 정의하면 된다.

- 데이터가 우리가 흔히 보는 자료형태가 아닐 때

- 예) 집합

- 커널을 이용해서 마치 일반적인 데이터인 것처럼 다룰 수 있음

- 예) 커널을 교집합의 크기로 정의

커널 + 커널 기호 유사성 가능하다.
RBF + customized.