

SVM. (Support Vector Machine). 191008

전통적인 분류방식, 즉

분류 - 여러 데이터에 대해 feature 을 기준으로 Boundary 를 나누어
데이터에 대해 class. 레이블을 주어 구분하는 것

위분류의 Boundary 는. decision function 이라고 일컫는다.

선형 (linear)
비선형 (non-linear)

- 3차원 이상의
비선형 Decision function

⇒ Hyperplane.

SVM 알고리즘.

W : Hyperplane Vector.

$$Y = \text{Sign}(W^T X + b) = f(x)$$

- $Y_i = +1$ when $W^T X + b > 1$
- $Y_i = -1$ when $W^T X + b < -1$

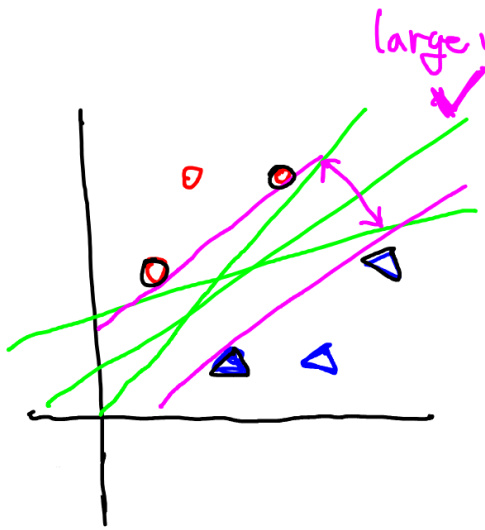
b : Margin

Categorical Data : 몇개의 범주로 정의된 데이터

ex) 남여, 고생계/사파 ...

Data Embedding

= 위 범주형 데이터를 벡터로 표현한 것.



→ 이는 Decision function 이 중요하다?

판단기준 : margin.

margin이 넓은 것이 좋은 Decision

function. 이를 위한 Test는 위 그림에 많은 양의 데이터를 TEST 해보면 된다.

Support Vector. — margin 크기의 Data,,