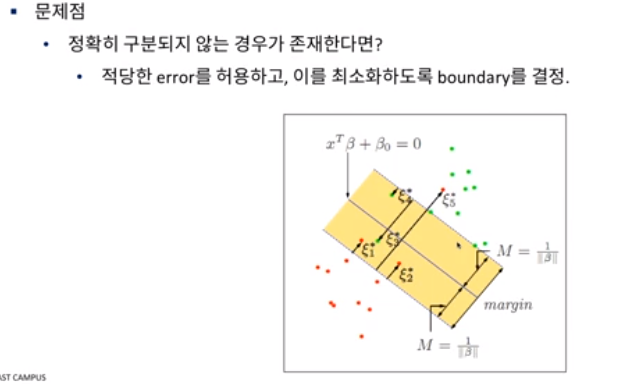
**<SVM (support vector machine)>**

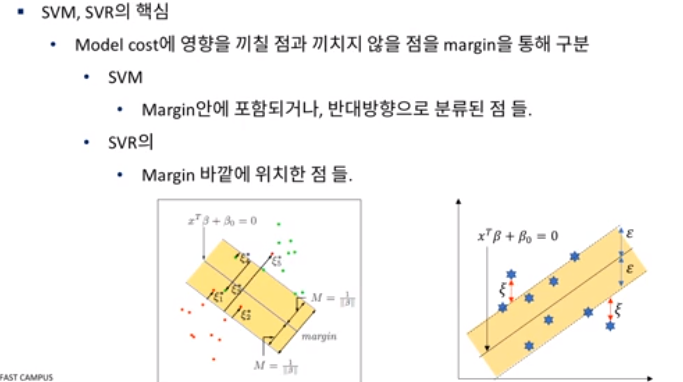
데이터의 분포가 정규분포도 아니고 그렇다할 생각나는 분포가 없다. : LDA를 적용하기 힘들다

무슨기준으로 나누는가? 최대한 겹치지 않으면서 샘플사이로 boundary를 찾는데 체계적으로 찾는 방법이 없을까? Margin을 최대화하는 boundary를 찾자.

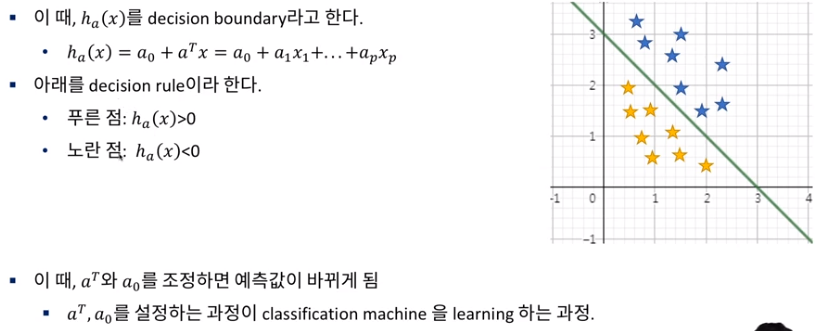


종속변수가 범주형변수 : SVM

종속변수가 연속형변수 : SVR ( support vector regression )



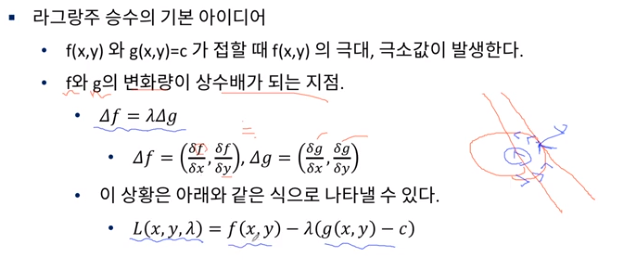
Cost가 높으면 버리고 낮으면 선택 , Model cost에 모든 점을 반영해서 구하는 게 아니라 영향을 끼치는 점(margin을 통해서 구분된)들을 통해 계산.



Decision boundary는 신경망에서도 사용된다. 추후 언급 (Tree구조에서도 언급)

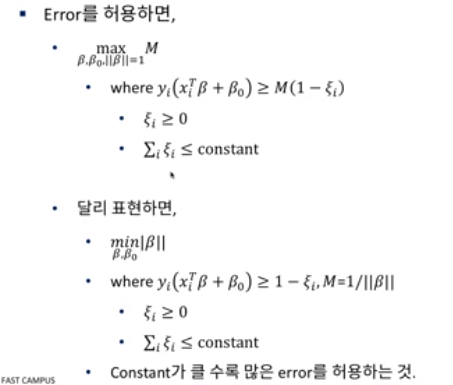
따라서 SVM은 decision boundary의 계수(weight)를 구하는 과정

**<라그랑주 승수>**

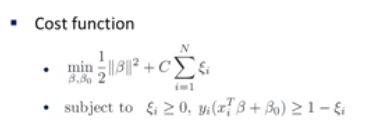


f(x, y)를 최대화 하는 것은 L(x, y, )를 최대화 하는 것과 같다.

Constant가 클수록 많은 error를 허용하는 것



목적 : Max M을 하면서 합이 작도록한다.

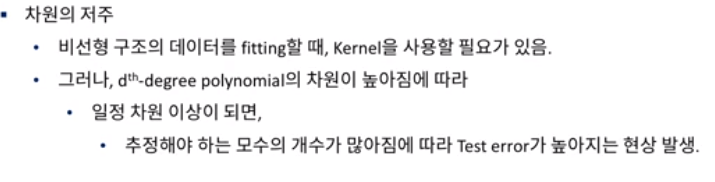
margin을 최대화한다. 베타를 최소화환다.

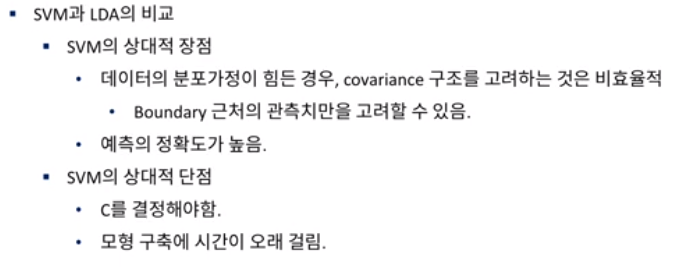
Margin = 1/절댓값 베타

라그랑주 성질을 이용해 support vector를 사용한다. (강의 참고 .. 너무어려워ㅠ)

**SVM with kernel**

Ex) 2차원의 데이터를 3차원으로 차원을 증가시켜 SVM을 구한다.





C : error의 상수 C가 작으면 error가 많아도 된다/ C가 무한대면 error가 전혀없도록

시그마가 크다면 차이의 정도가 완화된다 : 모양이 좀더 간단해진다.

**<SVR support vector regression>**

입력변수가 연속형 변수일 때 사용한다. Margin밖에 있는 점들의 에러가 최소가 되도록 하는형태로 동작한다.