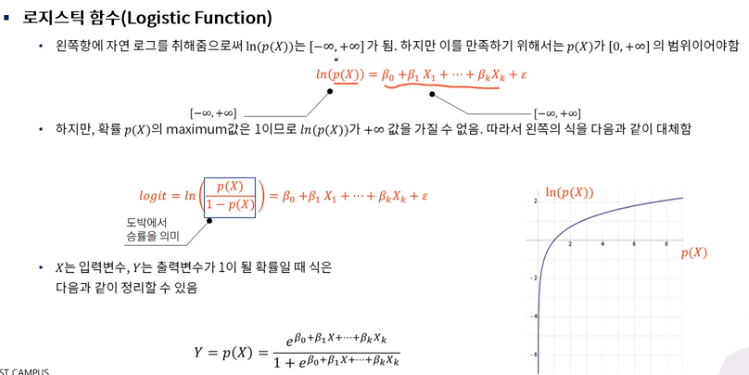
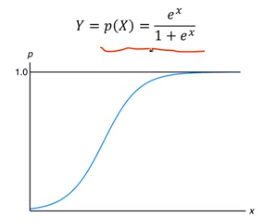
**< 로지스틱 회귀분석 > y가 실수형이 아닌 class를 가질 때 판단하는 분석기법**

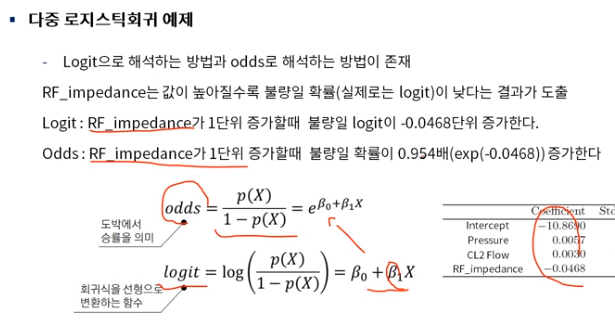
두 개의 카테고리를 가지는 형태의 출력변수. 확률값을 갖는다고 생각하고 식을 세운다.

선형회귀분석에서는 범위가 -무한 ~ +무한으로 범위가 나오는데 확률값은 0부터 1로 축소해줘야한다.





로지스틱 함수 또는 시그모이드 threshhold에 따라 클래스를 구별한다. Ex) 0.5이상은 정상 이하는 비정상



Logit이 -0.0468단위 줄어든다는게 잘 안와닿으면 불량률이 줄어든다 즉, 음의방향만 봐도 되지만 좀더 자연스럽게 표현하려면 exp를 취한 값의 배수만큼 증가한다 즉, 0.95배 증가한다( 불량률은 기존 1보다 작아짐 ) 이라고 생각한다.

**< 회귀계수 축소법 >**

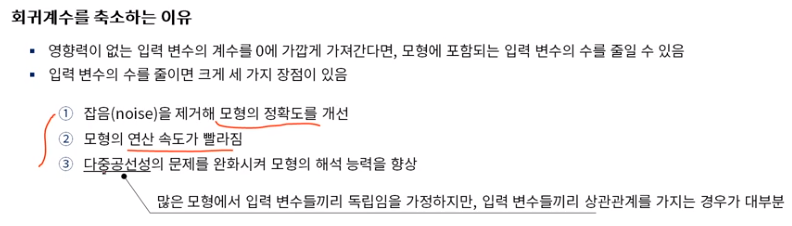
분석하기 좋은 데이터

* X,y사이에 관계가 분명
* X(독립변수)가 너무 많지않아야하고 ( 과적합방지, 학습시간증가 )
* 독립변수끼리는 상관관계가 적아야함
* 노이즈가 없는 깨끗한 데이터

필드에는 이런 양질의 데이터가 거의 없음, y값이 있는 데이터도 별로 없음. 따라서 variable selection을 진행해서 좋은 변수들을 뽑아내야한다.

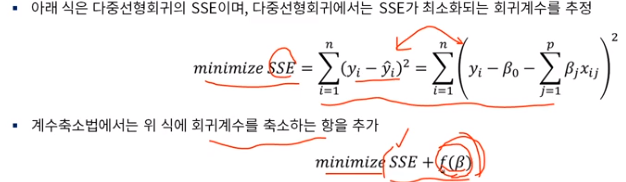
좋은 변수란?

Y의 변동성을 잘 설명하면서 x들끼리는 상관관계가 없는 변수들이 좋은 변수( 다중공선성을 제거해야한다 ) -> 따라서 회귀계수를 축소시켜 종속변수에 크게 영향을 끼치지 않는 변수를 제거하거나 줄이고 독립변수끼리 중복성이 있는 것들도 마찬가지로 처리



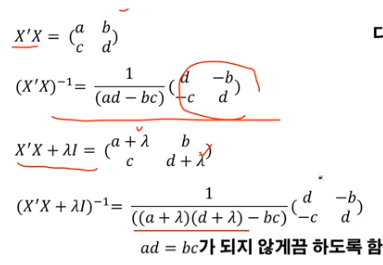
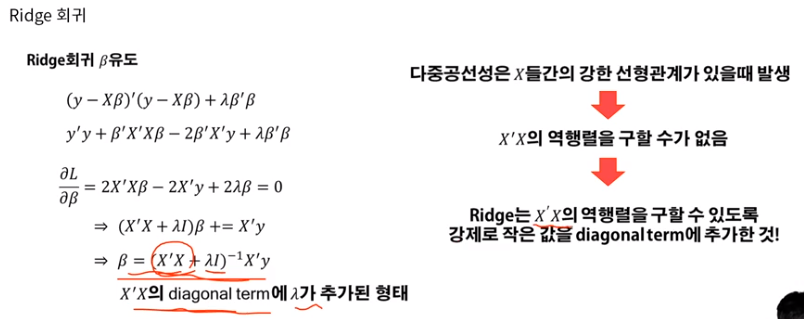
계수축소법의 종류

* Ridge회귀, Lasso회귀, Elastic-Net회귀



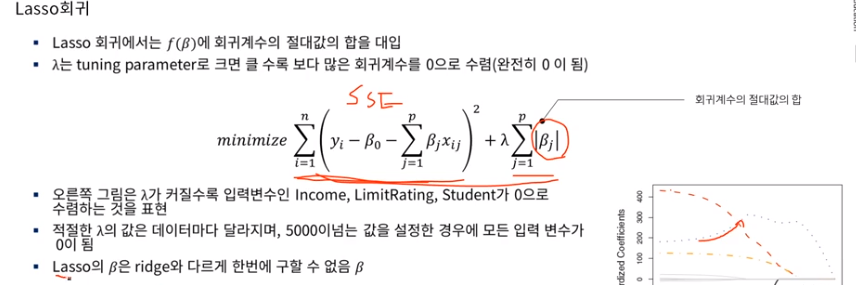
* 각 회귀분석이 다른점은 f(베타)부분의 회귀계수를 축소하는 항만 다르다.

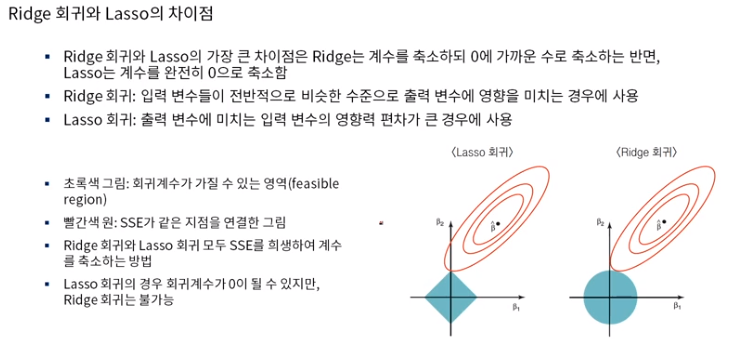
**Ridge**



**베타는 어떻게 구하는가? 참고 :** 

**Lasso**

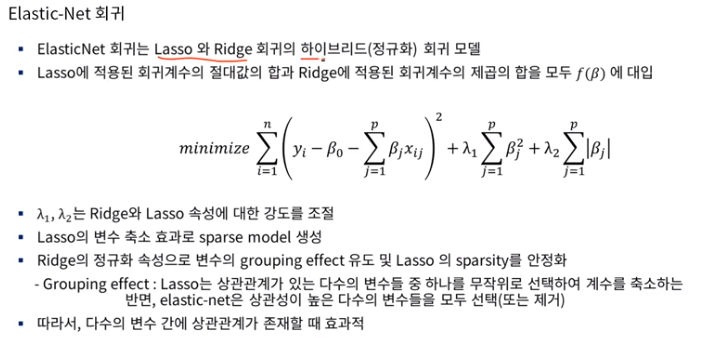




Lasso에서는 변수를 selection할수 있다(회귀계수를 아예 0으로 축소하기 때문에 필요없어지므로)

Lasso는 절댓값보다 작으므로 마름모꼴에 닿을 수 있다. / Ridge는 제곱합의 형태이다보니 원의 형태에 가까이간다. 닿지는 못한다.

**Elastic-Net**



**PCA부분 미수강 추후 공부필요**