

In []:

05.02 회귀분석의 기하학

1) 회귀 벡터공간

- $\hat{y} = Xw$
- \hat{y} 은 x 의 각 열을 기저벡터로 하는 벡터공간 내 존재 (span of X)
- $\hat{y} = y$ 를 x 가 이루는 벡터공간에 투영한 벡터 = Hy (H = 투영행렬)
- $e = My$ (M = 잔차행렬) (1page)

2) 잔차행렬과 투영행렬 (=햇행렬, 영향도행렬)

- y 를 각각 잔차와 예측값으로 변환하는 행렬
- 잔차행렬과 투영행렬의 성질 (4page) (꼭 암기)

- y 벡터의 제곱합 = 잔차 벡터 e 제곱합 + 추정치 벡터 \hat{y} 제곱합 (6page)
(증명해보기)

05.01 확률론적 선형 회귀모형

4) 잔차의 분포

- 확률론적 선형회귀모형 : " $\text{잔차} = e = y - w \cdot Tx$ 도 정규분포따름"
- 확률론적 선형회귀모형에서는 잔차와 잡음이 다른 개념이다.
- 잡음이 정규분포이면 (확률론적 선형 회귀모형 하), 잔차도 정규분포를 따른다. (7page 증명!)
- 잔차는 잡음의 선형변환. 따라서, 잡음의 가정들이 잔차에도 적용됨
ex) 잔차의 기대값 = 0

5) 회귀계수의 표준오차

- 다시 처음으로, 확률론적 선형회귀모형 쓰는 이유는? = "우리가 예측한 것의 오차는?" 에 답하기 위해
- "우리가 예측한 것의 오차" = 회귀계수의 표준오차(se, Standard Error)
- 증명은 나중에 시간 되면..

[se] (9page 필기)

정규화된 모수(w) 오차 ==>> 표준스튜던트t분포를 따름 (자유도 N-K)

*N = 표본 데이터 수, K = 가중치 갯수(0~K

-1 까지 가중치)