코릴레이션correlation

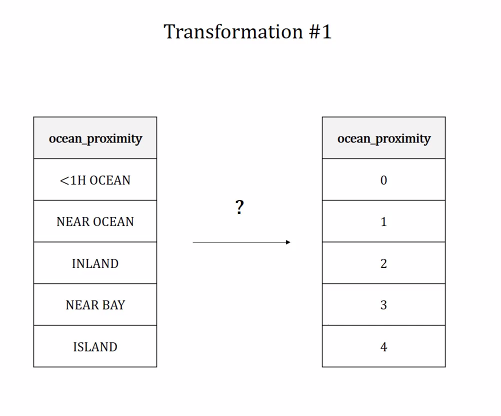
x와 y간의 관계?

부호가 양수일 때

부호가 음수일 때

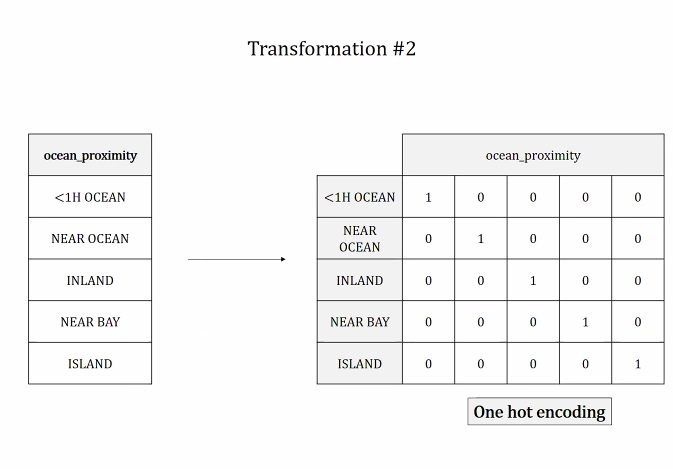
correlation의 대각선 값 = 1 자기자신이니까

왜곡

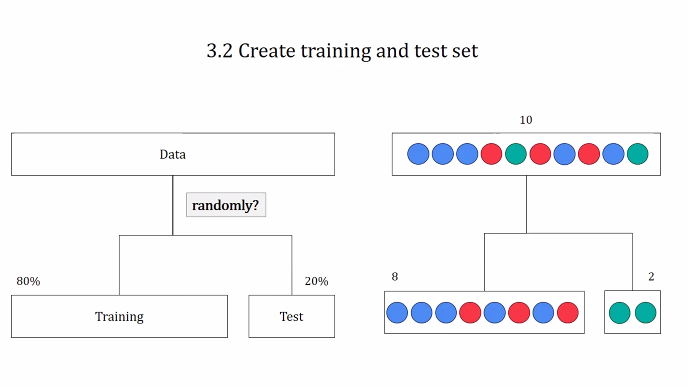


위는 스칼라로 같은 방향에 다른 값을 갖게 됨으로 왜곡이 생김(=scaling, 스케일링)

각 벡터는 방향성을 갖는다. 서로 다른 방향을 갖고 있다를 의미



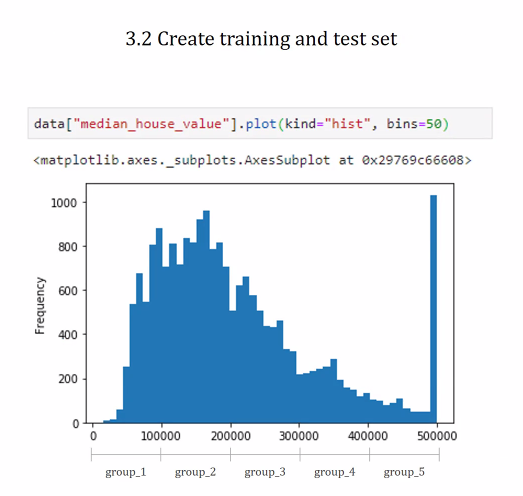
Training / Test 데이터 나눌 때 유의해야 할 점?



Random 하게 뽑지만 Training Data에 분류 빨 파만 학습시키고 초록색으로 테스트?오류

균등하게 학습을 시켜 주어야 한다.(즉 Stratified sampling 층별 추출법을 해야한다.)

그룹을 나누어서 1,2,3,4,5가 균등하게 들어가게 샘플링



Missing value를 training data와 test date set을 나누는 이유는

미리 데이터를 삭제하거나 중앙 값을 넣게되는 경우 test data를 건들게 되기 때문에