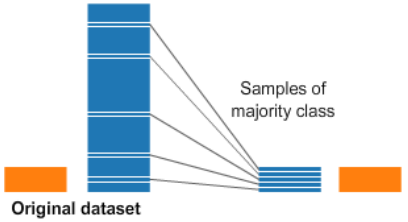
# [참고] 리샘플링 알고리즘 비교

## ❶ Random Oversampling

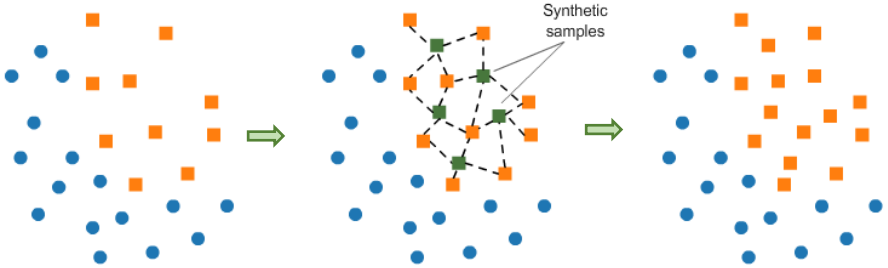

(특징) 소수 클래스(minority)을 랜덤 과대 표집
(장점) 원 데이터의 특성을 잃지 않음
(단점) 소수 클래스의 대표성을 담보할 수 없음, 오버피팅 가능성

## ✓❷ Random Undersampling



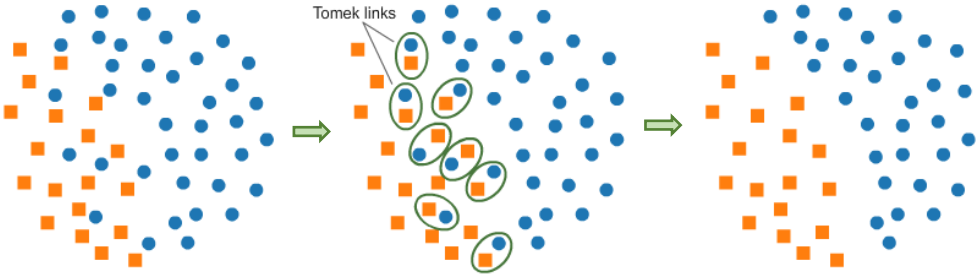

(특징) 다수 클래스(majority)를 랜덤 과소 표집
(장점) 빅데이터에 적합, 학습 시간을 줄일 수 있음
(단점) 샘플의 크기가 현저히 감소함

## ✓❸ SMOTE




(특징) 소수 클래스의 통계적 특성을 합성하여 데이터 포인트 추가
(장점) 소수 클래스에서 있을 수 있는 noise 제거
(단점) 고차원 데이터에서 효과적이지 않을 수 있음

## ❹ Tomek Links




(특징) 소수 클래스와 근접한 다수 클래스의 데이터 포인트를 삭제
(장점) 클래스간 거리를 넓혀 분류가 용이해짐
(단점) 기존 원 데이터에서 판별 특성이 명확해야 함