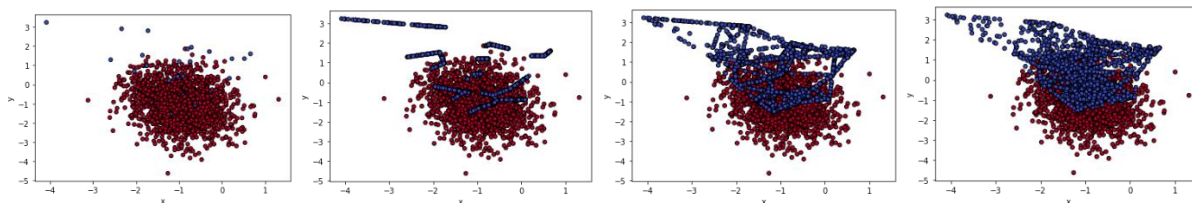
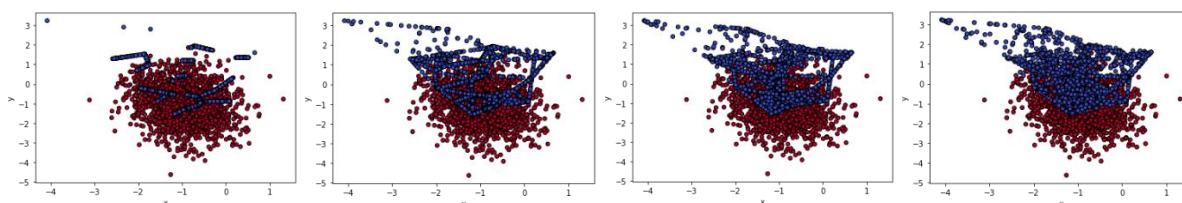


1. 2개의 클래스를 가진 2000개의 불균형 자료(1% : 99%)이다.

(1) SMOTE를 적용하면 k=1일때 선의 형태로 합성관측치가 생겨나 k는 2이상이어야 한다는 것을 알 수 있다. k가 5, 15로 늘어남에 따라 class의 경계가 다수클래스의 데이터에 치중되지 않고 class 경계선이 아래쪽으로 내려온다. k가 늘어날수록 합성관측치가 더 넓은 범위에 분포된다.



(2) ADASYN을 적용했을 때 k=1일 때 다수클래스가 주변에 많은 관측치에서 선의 형태로 합성 관측치가 생겨난다. 또한 SMOTE와 마찬가지로 k가 5, 8, 15로 늘어남에 따라 class의 경계가 다수클래스의 데이터에 치중되지 않고 class 경계선이 아래로 내려오고 더 넓은 범위에 합성관측치가 생성되어 분류에 영향을 준다.



2. chi2, f_classif, mutual_info_classif를 적용해 20개의 특성변수를 선택하였다.

```
False False False False False True True False False False False False False False False
False False False False False True False False False False True False False False False
False False False True True True False False False False False True True False False
False False True False True False True False False False True True False True False True
False True True True True False True False False False True True True False True False
False False False False False False True False False False False False False False True
False False True False False True True False False False False True True False False

False False True False False False False False
False False True False False True False False
False False False False True True False False
False False True False True False True False
False True True False True False True False
False False True True True False True False
False False False False False True True False
False False True False False True False False
```

선택된 특성변수는 True로 나타난다. chi2, f_classif, mutual_info_classif 모두 대부분의 특성변수가 중앙에 위치한 변수에서 선택되었다. 따라서 변수선택이 잘 되었다. 위와 같이 변수를 8 X 8 로 나타내고 3열, 4열, 5열, 그리고 6열에 위치한 변수를 중앙에 위치한 변수라고 할 때, 선택된 특성 변수의 개수는 chi2에서는 13개, f_classif에서 15개, mutual_info_classif에서 15개가 있었다.

F와 chi2에서 선택된 특성변수는 15개가 서로 동일하다. 즉 Large sample 일때 F가 chi2로 수렴해 chi2와 F는 통계적으로 같은 성격을 가진다는 것을 알 수 있다.