

Machine Learning

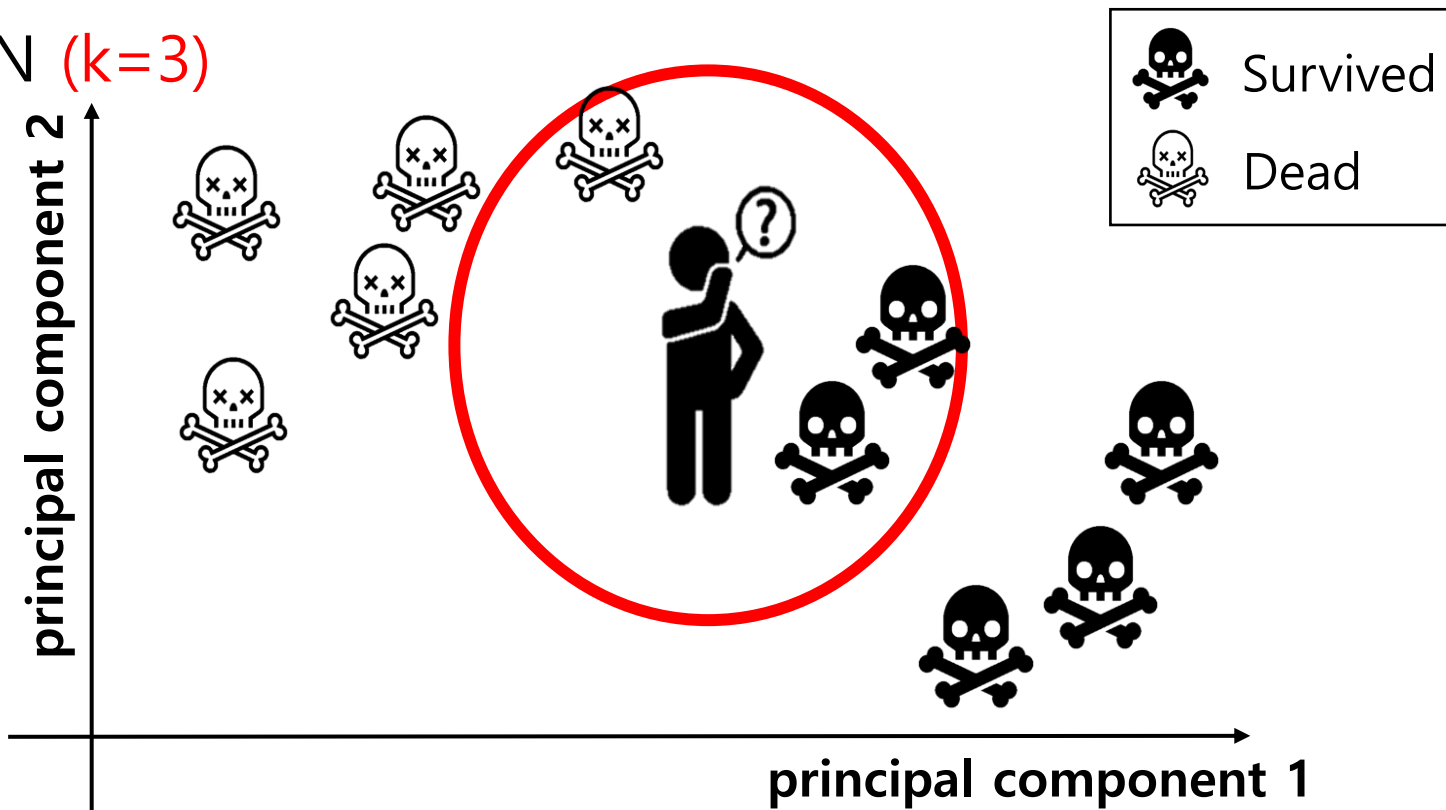
K-Nearest Neighbors(KNN)

k-최근접 이웃 알고리즘

- 새로운 데이터 포인트와 가장 가까운 훈련 데이터셋의 데이터 포인트를 찾아 예측
- k 값에 따라 가까운 이웃의 수가 결정
(단, k가 짝수이거나 클래스 개수의 배수는 지양)
- 분류(다수결)와 회귀(평균)에 모두 사용 가능

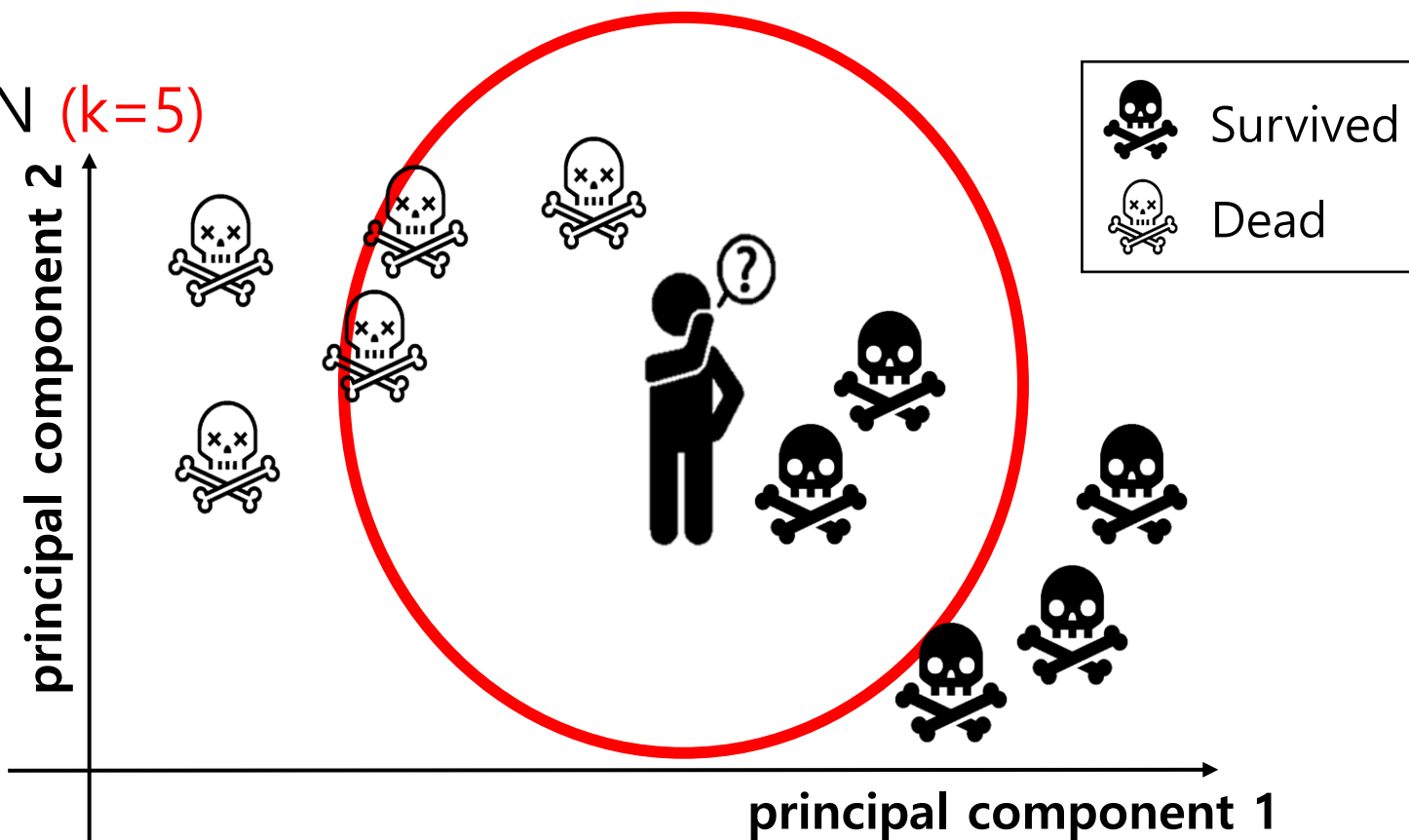
K-Nearest Neighbors (KNN)

k-최근접 이웃 알고리즘

KNN ($k=3$)

K-Nearest Neighbors (KNN)

k-최근접 이웃 알고리즘

KNN ($k=5$)

장단점

- 이해하기 매우 쉬운 모델
- 훈련 데이터 세트가 크면(특성, 샘플의 수) 예측이 느려진다.
- 수백 개 이상의 많은 특성을 가진 데이터 세트와
특성 값 대부분이 0인 희소(sparse)한 데이터 세트에는
잘 동작하지 않는다.
- 거리를 측정하기 때문에 scaling 혹은 정규화 필요