

Machine Learning

An Introduction to Statistical Learning

- Chapter 03 with Bootstrap method

허성욱

acto8290@gmail.com

<https://github.com/AnalystH/>

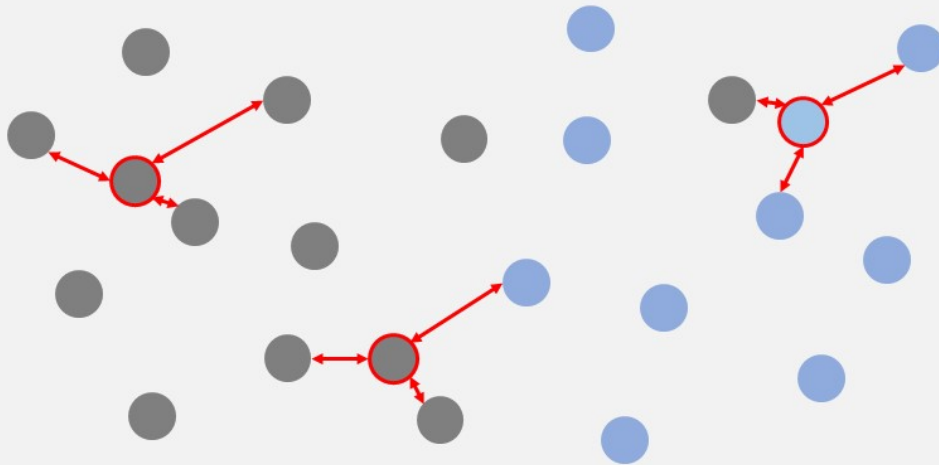
✓ K-Nearest Neighbor(K-NN)

➤ **K-NN** : 기존 데이터 중 가장 유사한 k 개의 데이터를 이용해서 값을 예측하는 방법.

- 예측 대상 :
 - 수치형 : k개 데이터의 평균(또는 가중 평균)
 - 명목형 : k개 데이터 중 많이 나온 분류 항목 선택 (또는 가중 값 이용 분류 선택)
- 유사도 측정 : euclidean distance (2-norm) 이용.
- 변수간의 값 차이가 클 경우 정규화를 한 후 분석을 해야 한다.
 - $(\text{value} - \text{min}) / (\text{max} - \text{min})$

✓ K-Nearest Neighbor(K-NN)

K-Nearest Neighbor



- $k=3$ 일 때 새로운 점에 대하여 근처 3개의 점의 색깔에 따라 새로운 점의 색이 결정이 됨.

✓ K-Nearest Neighbor(K-NN)

➤ Advantages

- 비모수적 방법이기 때문에 어떤 분포든 상관 없음
- 쉽고 이해하기 직관적
- 샘플수가 많을 때 좋은 분류 법이다.

➤ Disadvantages

- 최적의 k를 선택하기가 어렵다.
- 데이터가 많을 때 분석 속도가 느릴 수 있음.
- 특정 분포를 가정하지 않기 때문에 샘플 수가 많이 있어야 정확도가 좋다.



경 청 해 주 셔 서
감사합니다