

## 2.2 잡음(noisy data) 제거

### ➔ 잡음(noise)

- 변수(속성)에서의 오류나 오차 값
- 오류나 오차에 의한 값의 경향성 훼손을 줄이기 위해서 데이터 평활화 기법(smoothing technique)을 적용

### ➔ 데이터 평활화 기법

- 구간화(Binning)

정렬된 데이터 값들을 몇 개의 빈(혹은 버킷)으로 분할하여 평활화하는 방법

- ① 평균값 평활화(smoothing by bin means)
- ② 중앙값 평활화(smoothing by bin medians)
- ③ 경계값 평활화(smoothing by bin boundaries)

- 구간화(Binning) 방식

- ① 동일 너비 방식
- ② 동일 높이 방식

## 2.2 잡음(noisy data) 제거

### 평활화 예제 [동일 너비(범위) 방식]



## 2.2 잡음(noisy data) 제거

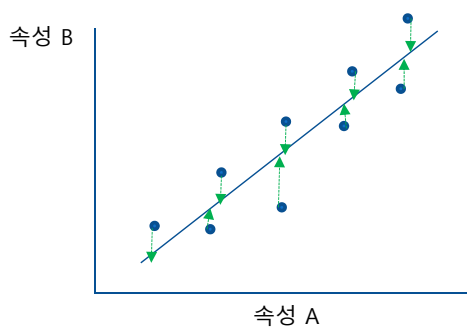
평활화 예제 [동일 높이(개수) 방식]



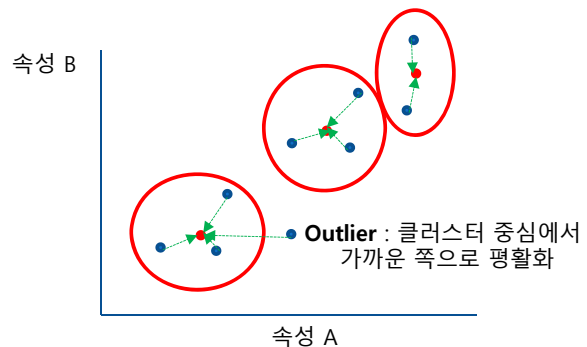
## 2.2 잡음(noisy data) 제거

### 데이터 평활화 기법

- 회귀(Regression)
  - 회귀 함수에 의한 데이터 평활화 기법
    - ✓ 선형회귀분석 : 하나의 속성값으로 다른 하나의 속성값 예측
    - ✓ 다중회귀분석 : 두 개 이상의 속성값으로 다른 속성값을 예측
- 군집화(Clustering)
  - 유사한 값들끼리 그룹화하는 기법 (outlier는 평활화 대상)



< 회귀에 의한 평활화 >



< 군집화에 의한 평활화 >