

#4 AdaBoost



11기 Java 김민혁

목차

4.1 의사결정나무

4.2 Ensemble

4.3 AdaBoost



Decision Tree

- Python

```
def 두 집단 간 이질성이 크게끔 나누자!():  
    if 분기되지 않는다:  
        return  
    두 집단 간 이질성이 크게끔 나누자!()
```

- Java

```
static ~~ 두 집단 간 이질성이 크게끔 나누자!(){  
    if (분기되지 않는다){  
        return;  
    }  
    두 집단 간 이질성이 크게끔 나누자!();  
}
```



Decision Tree

- Pros
 - 구조가 단순, 수학적 가정 불필요.
- Cons
 - 경계선 근처 값이 오차가 클 수 있음 (비연속성).
 - 예측변수의 효과 파악이 어려움.

Ensemble

- 단순 Decision Tree는 Not MZ (엠제트)

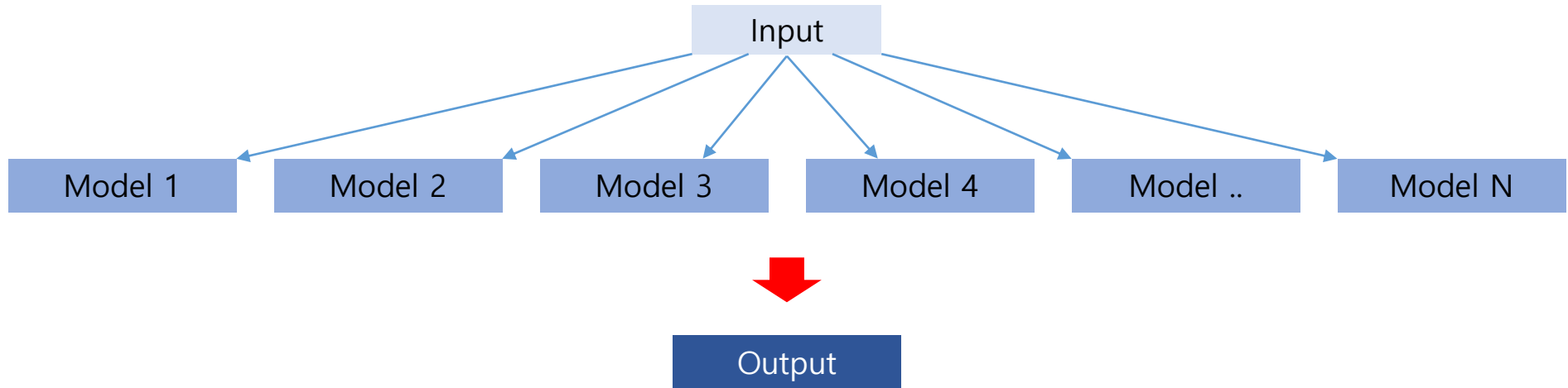
➡ 여러 개를 쌓아보자!

- Bagging
- Boosting



Ensemble

- Bagging



- Categorical Data: Voting
- Continuous Data: Average

Ensemble

- Boosting



- Bagging 보다 성능이 좋음
- 비교적 속도가 느리고, 과적합 가능성 ↑

AdaBoost



Yoav Freund

Effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules of thumb..

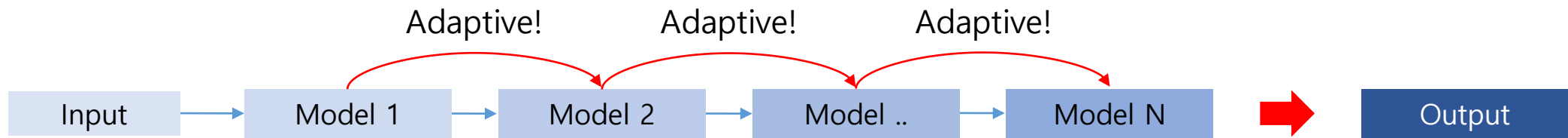
따봉도치야 고마워! (다소 의역)



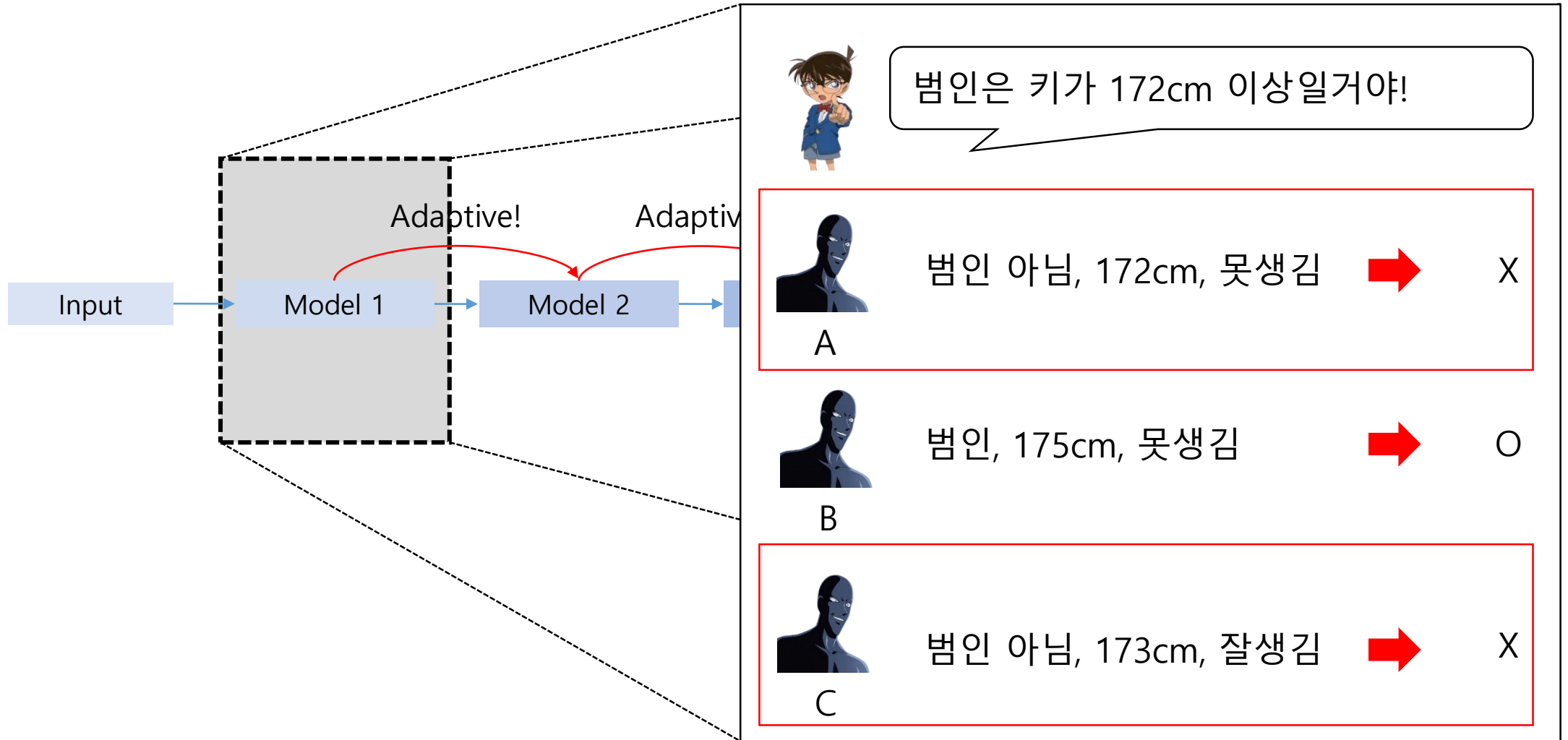
AdaBoost



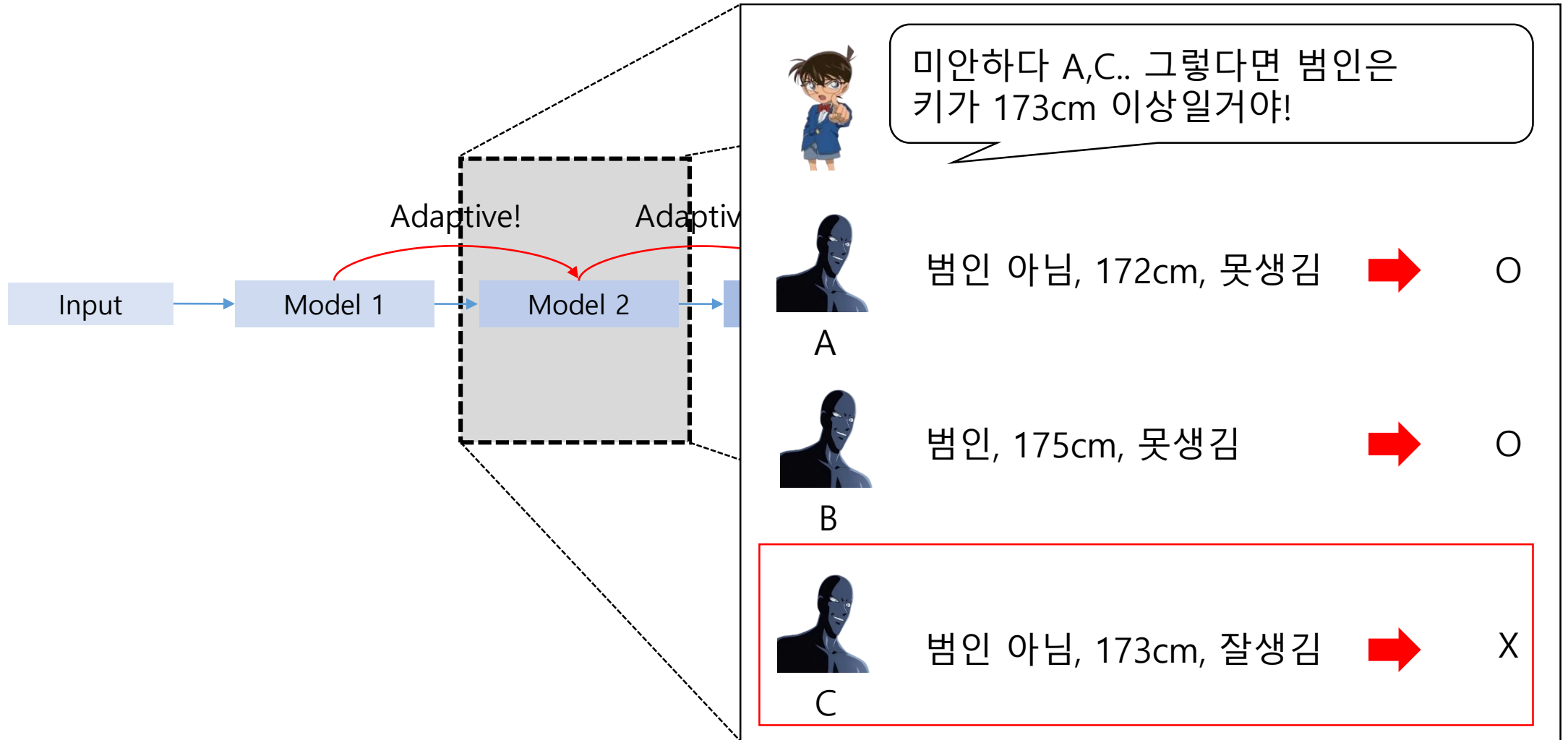
AdaBoost



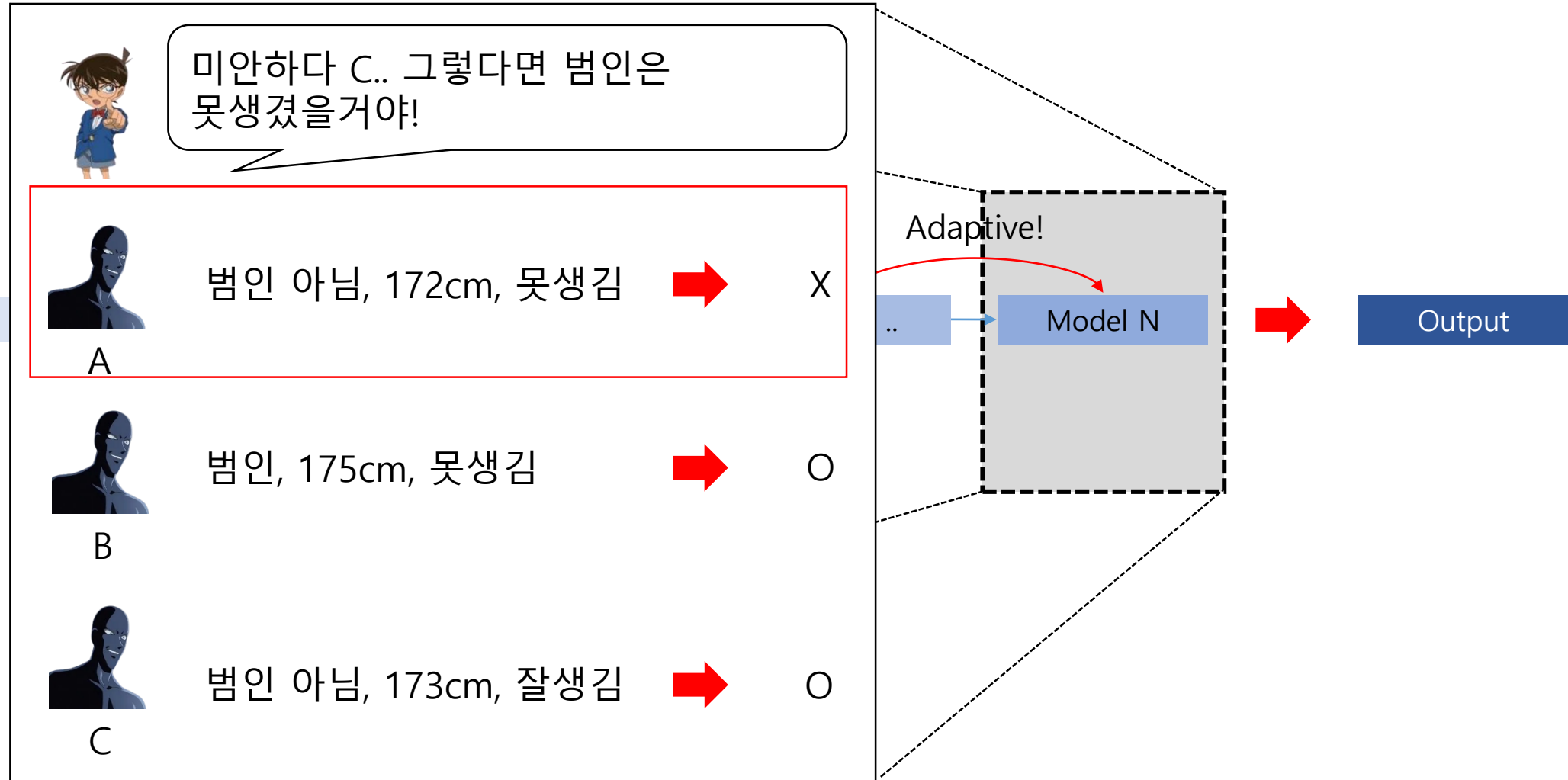
AdaBoost



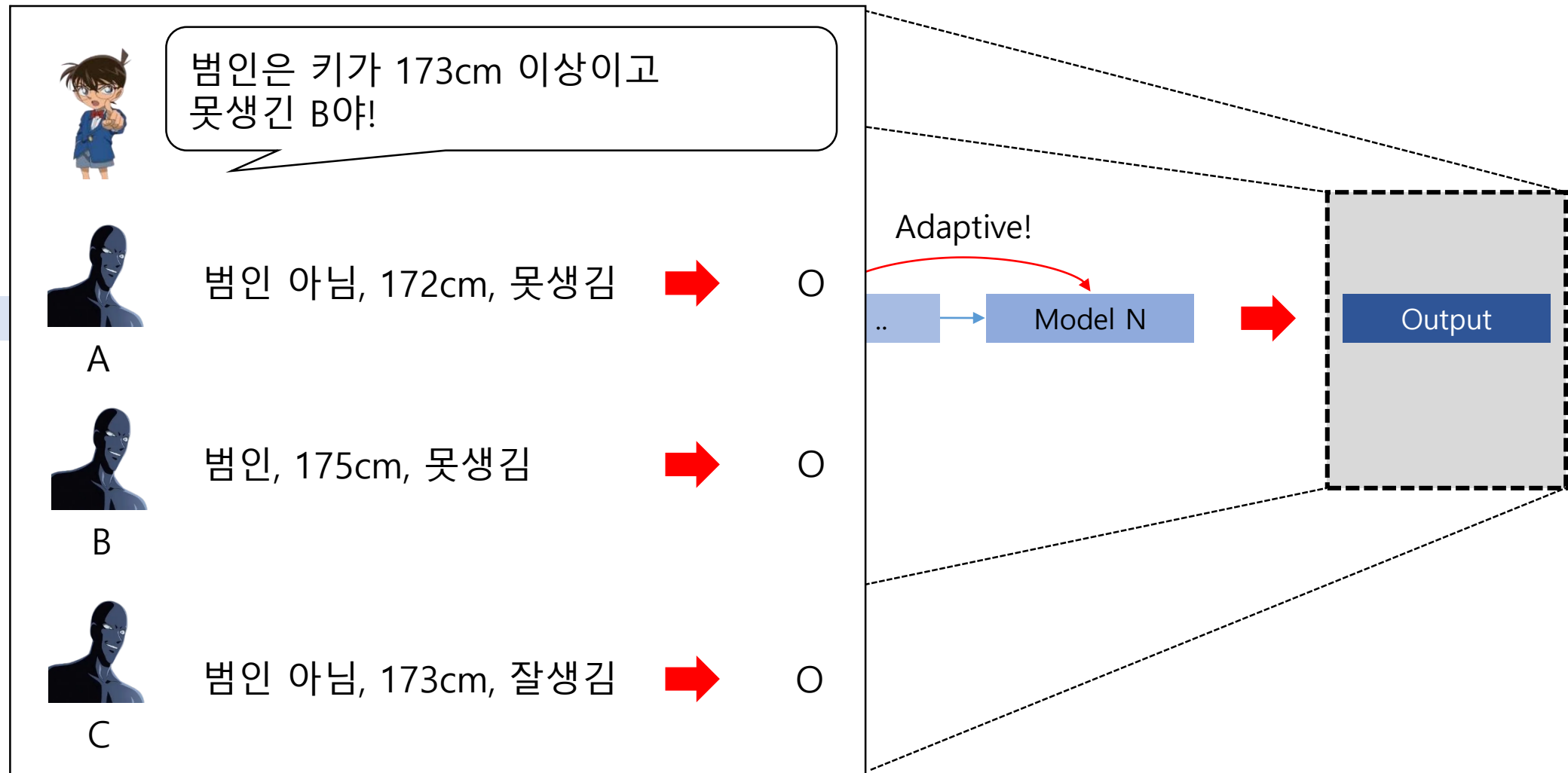
AdaBoost



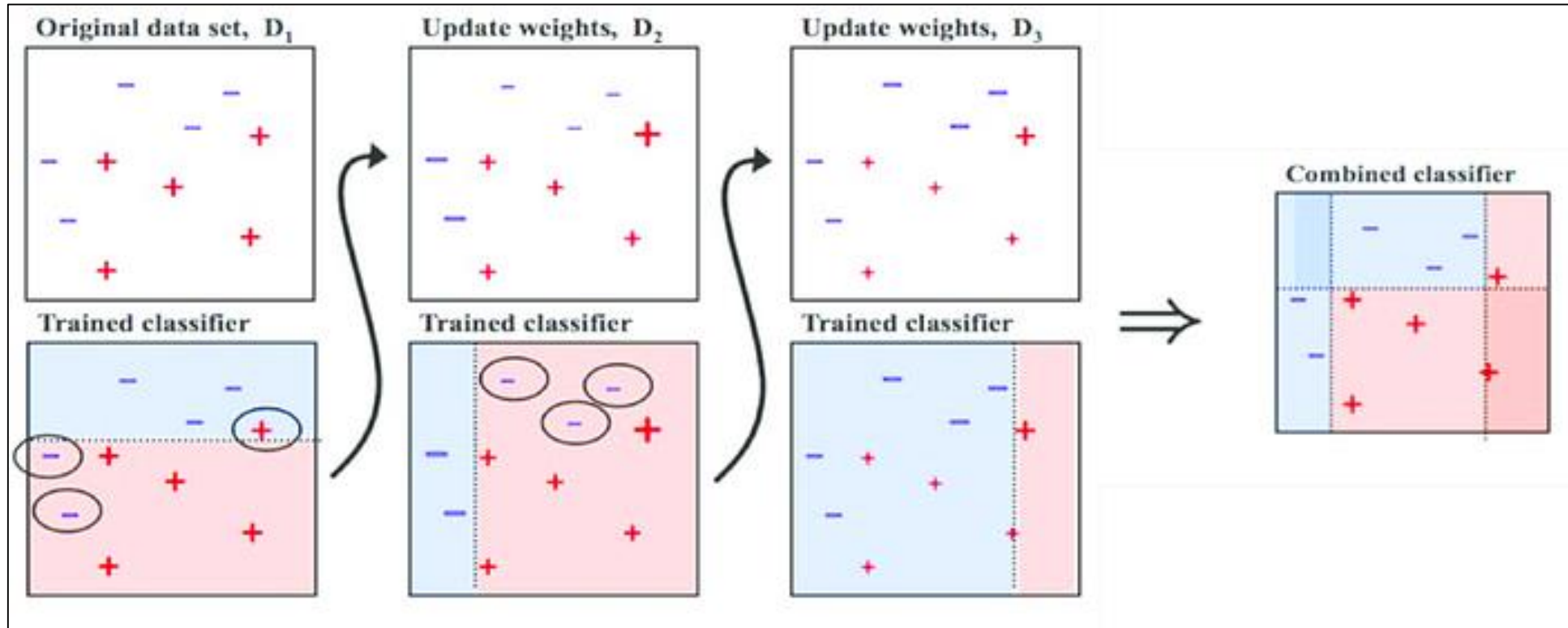
AdaBoost



AdaBoost



AdaBoost



AdaBoost

- Pros
 - 일반적인 Boosting보다 과적합의 영향을 덜 받음.
- Cons
 - 이상치에 민감함.
 - 해석이 어려움.

AdaBoost

- Hyper Parameter
 - base_estimators
 - 학습에 사용하는 알고리즘
 - default : None → DecisionTreeClassifier(max_depth = 1) 이 적용
 - n_estimators
 - 생성할 약한 학습기의 개수를 지정
 - default : 50
 - learning_rate
 - 약한 학습기가 순차적으로 오류값을 보정해나갈 때 적용하는 계수
 - default : 1



감사합니다.

