

Gradient Boost

- **Gradient Boost**

- GB: leaf로 부터 시작
 - Leaf: Target에 대한 초기 추정값(예: 평균, $\log(\text{odds ratio})$ 등)
 - Stump가 아닌 Tree를 생성: 각 tree는 leaf가 8~32개 크기 수준으로 생성

Target			
Height	Color	Gender	Weight
1.6	B	M	88
1.6	G	F	76
1.5	B	F	56
1.8	R	M	73
1.5	G	M	77
1.4	B	F	57

Gradient Boost

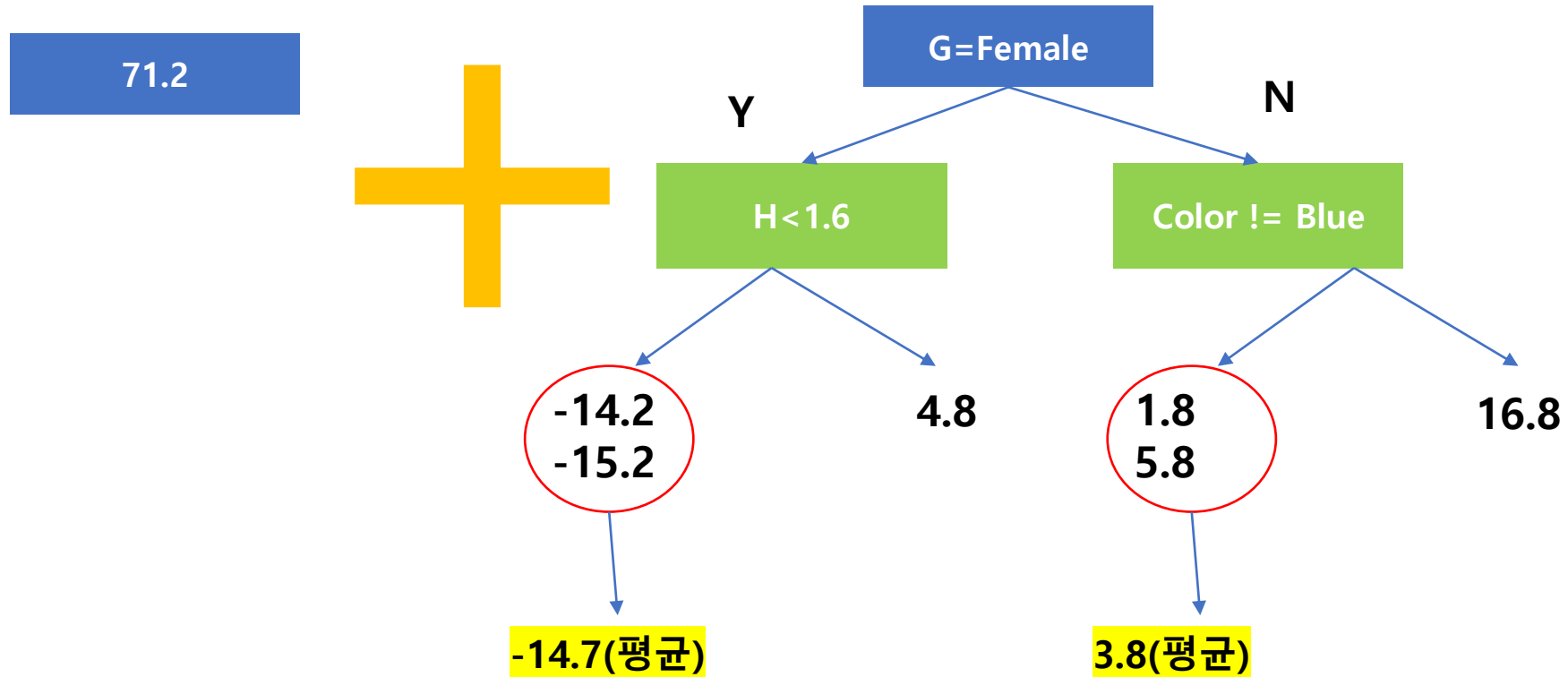
- **Gradient Boost, Step 1**
 - Leaf의 계산
 - Target인 Weight의 평균: 71.2
 - Residual을 계산: 실제값과 예측값의 차이(error)

같은 X변수들로
Residual에 대한 Tree

Height	Color	Gender	Weight	Residual
1.6	B	M	88	16.8
1.6	G	F	76	4.8
1.5	B	F	56	-15.2
1.8	R	M	73	1.8
1.5	G	M	77	5.8
1.4	B	F	57	-14.2

Gradient Boost

- Gradient Boost, Step 1
 - Leaf + **1st Tree**

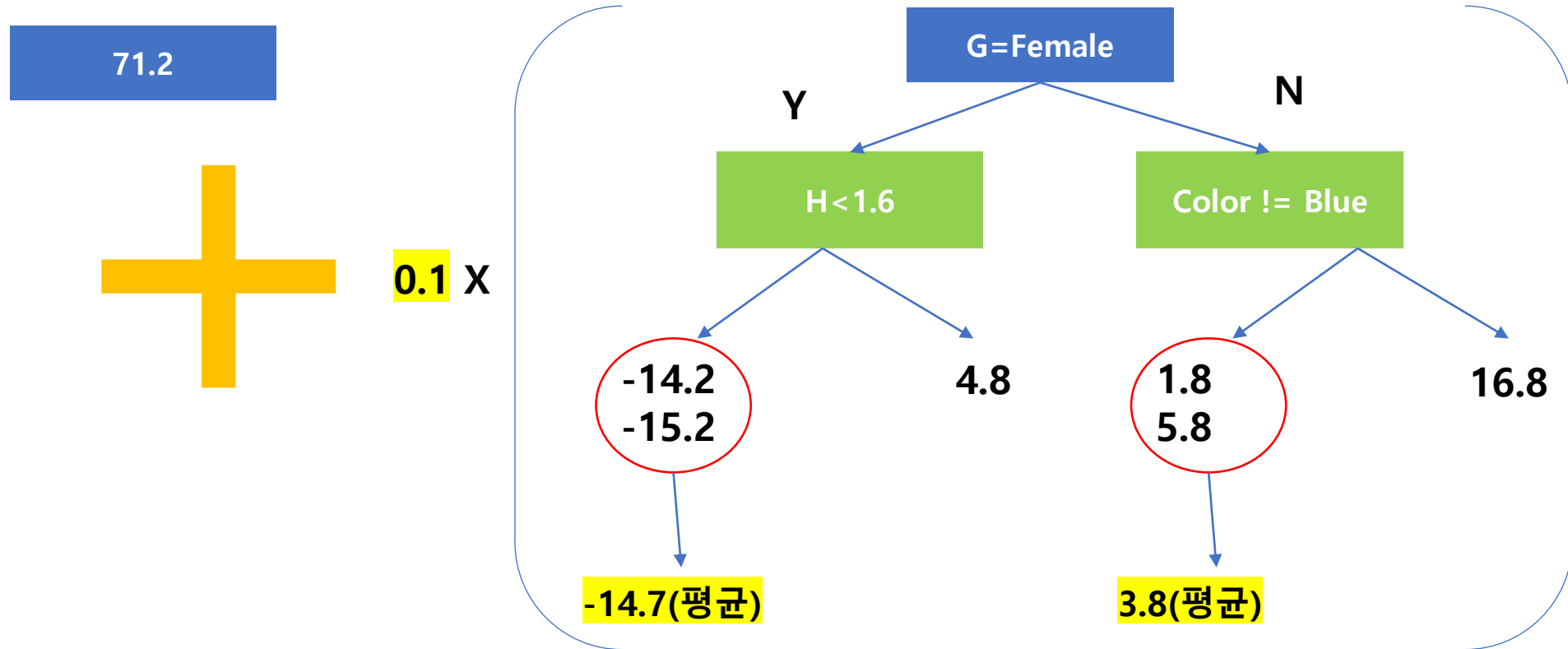


- Male, Blue인 경우 예측 예시:
 - $71.2 + 16.8 = 88$ (관측치와 동일하지만 과적합)
 - Bias는 작지만 Variance 큰 상태

Gradient Boost

- **Gradient Boost, Step 2**

- 과적합 방지, 학습속도 조절을 위한 학습율 도입
- Learning Rate: 0~1사이, 이 예에서는 0.1 사용

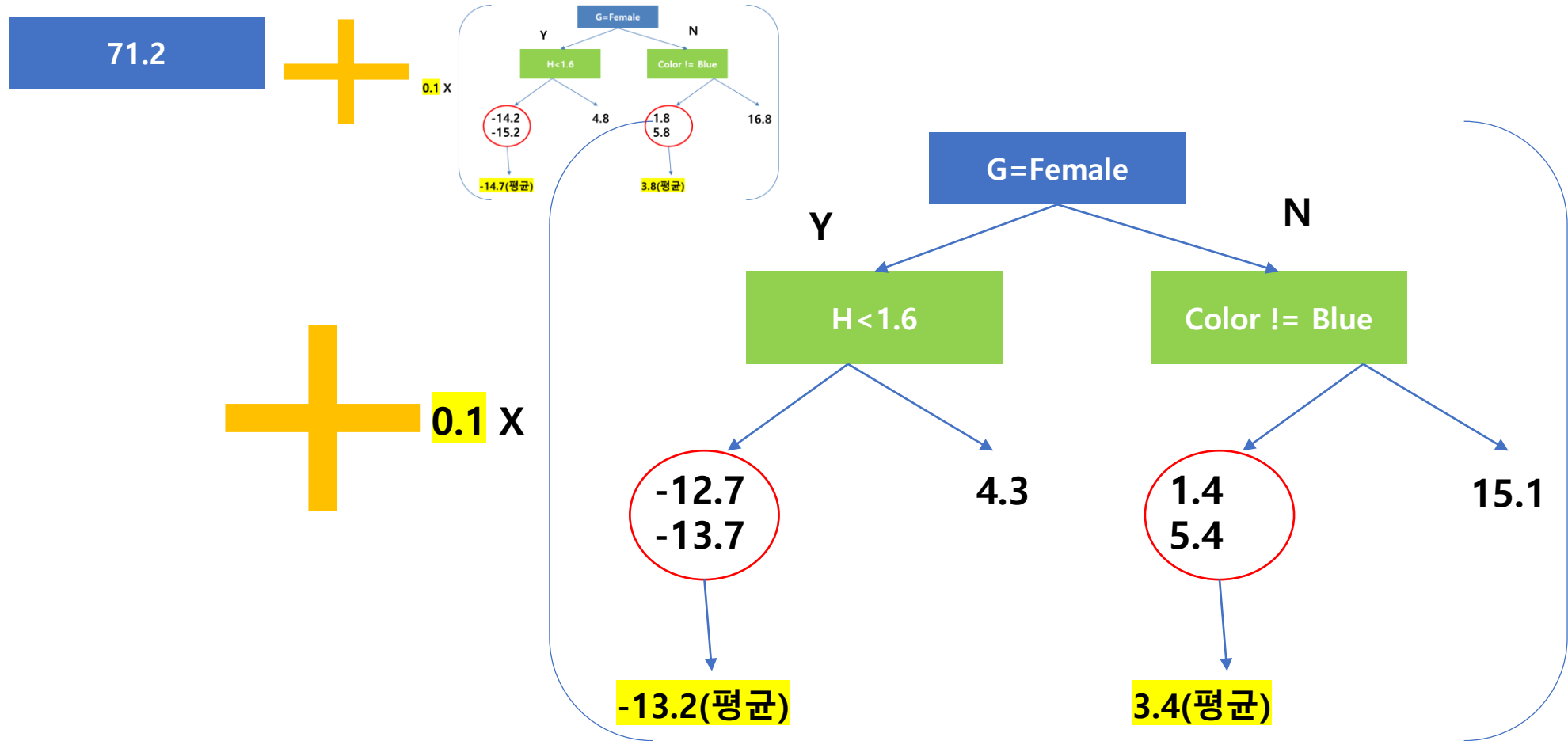


- **Male, Blue인 경우 예측 예시:** $71.2 + 0.1 \times 16.8 = 72.9$
 - 실제값에 가까워지지만, 그 정도가 조절됨 (Gradient의 개념)
 - Variance를 낮게 유지할 수 있음

Gradient Boost

- Gradient Boost, Step 3

- Learning Rate: 0~1사이, 이 예에서는 0.1 사용



- H=1.6, Male, Blue인 경우 예측 예시: $71.2 + 0.1 \times 16.8 + 0.1 \times 15.1 = 74.4$

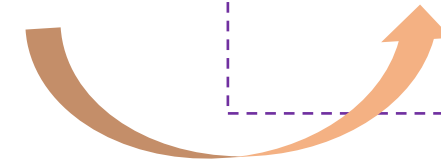
Gradient Boost

- **Gradient Boost, Step 3**

- 학습을 반영 예측값을 통한 두 번째 Residual 계산

같은 X변수들로 New
Residual에 대한 Tree

Height	Color	Gender	Weight	Residual	Residual(new)
1.6	B	M	88	16.8	15.1
1.6	G	F	76	4.8	4.3
1.5	B	F	56	-15.2	-13.7
1.8	R	M	73	1.8	1.4
1.5	G	M	77	5.8	5.4
1.4	B	F	57	-14.2	-12.7



Residual 크기 감소

Gradient Boost

- **Gradient Boost**
 - 위의 과정을 계속 반복
 - 정해진 iteration한도 까지 반복
 - 또는 이전 단계와 이후 단계의 Residual 차이가 없을 때까지 반복
 - 매 iteration에서의 Tree의 leaf는 8~32개 사이에서 생성
 - 매 iteration마다 다르게 생성
 - 1st tree: leaf 8개
 - 2nd tree: leaf 32개
 - 3rd tree: leaf 16개
 - ...