구간 계산으로
인공신경망 분석하기

고현수
hsgo@ropas.snu.ac.kr

2019. 05. 31. SnT

SMT Solver 기반 접근

- 하나도 빠짐이 없음(Complete)
- 한계: 뉴런 < 1,000 인 신경망
- 현재: 자율 주행에 이용되는 신경망의 크기: 뉴런 > 10,000

구간으로 분석하기

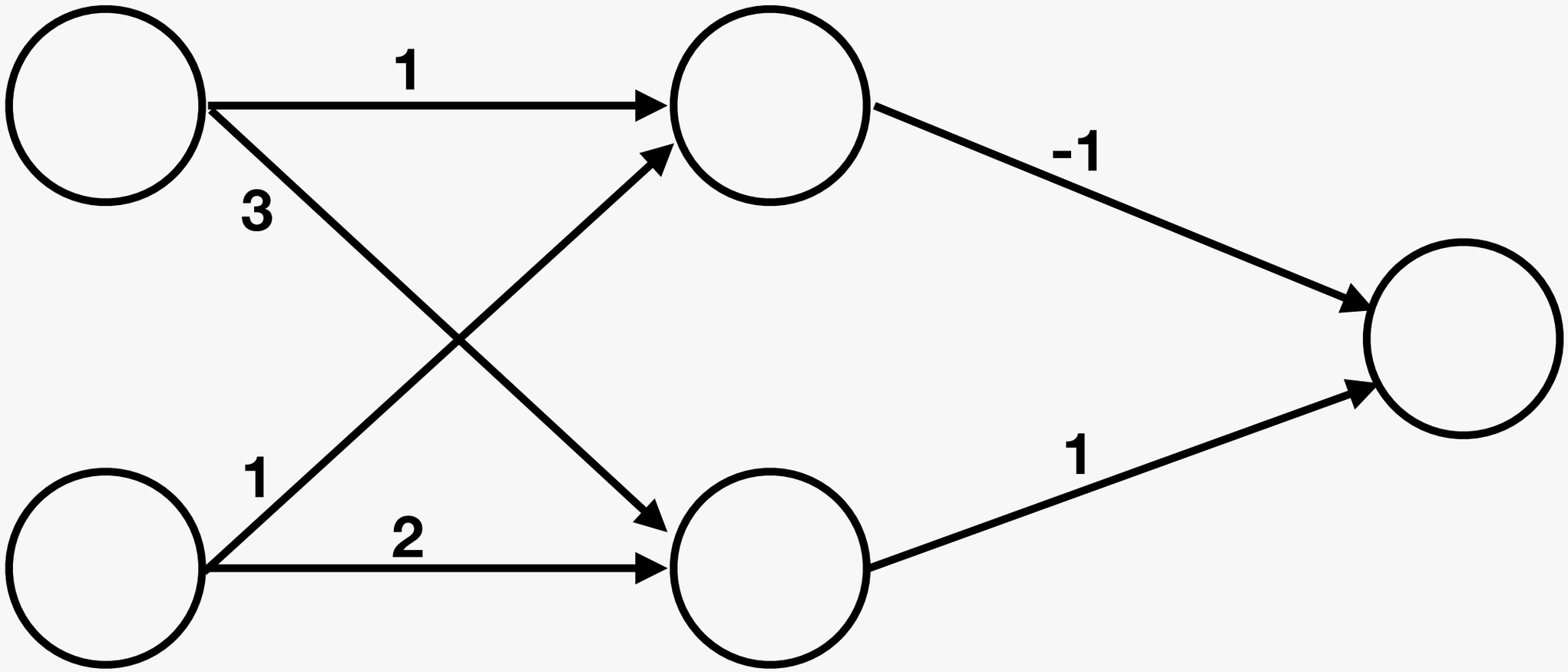
- 신경망은 연속적인 간단한 계산으로 이루어짐
  - 덧셈, 곱셈
  - 비선형 활성 함수(ReLU, sigmoid, tanh…)
- 구간으로 계산하기 쉬움

구간으로 분석하기

- 특정 구간의 입력에서 신경망이 올바르게 동작할 지
- 입력 구간에 대해 신경망의 결과 계산
- 구간 계산 결과는 신경망의 실제 결과를 포섭
- 구간 계산 결과 분석

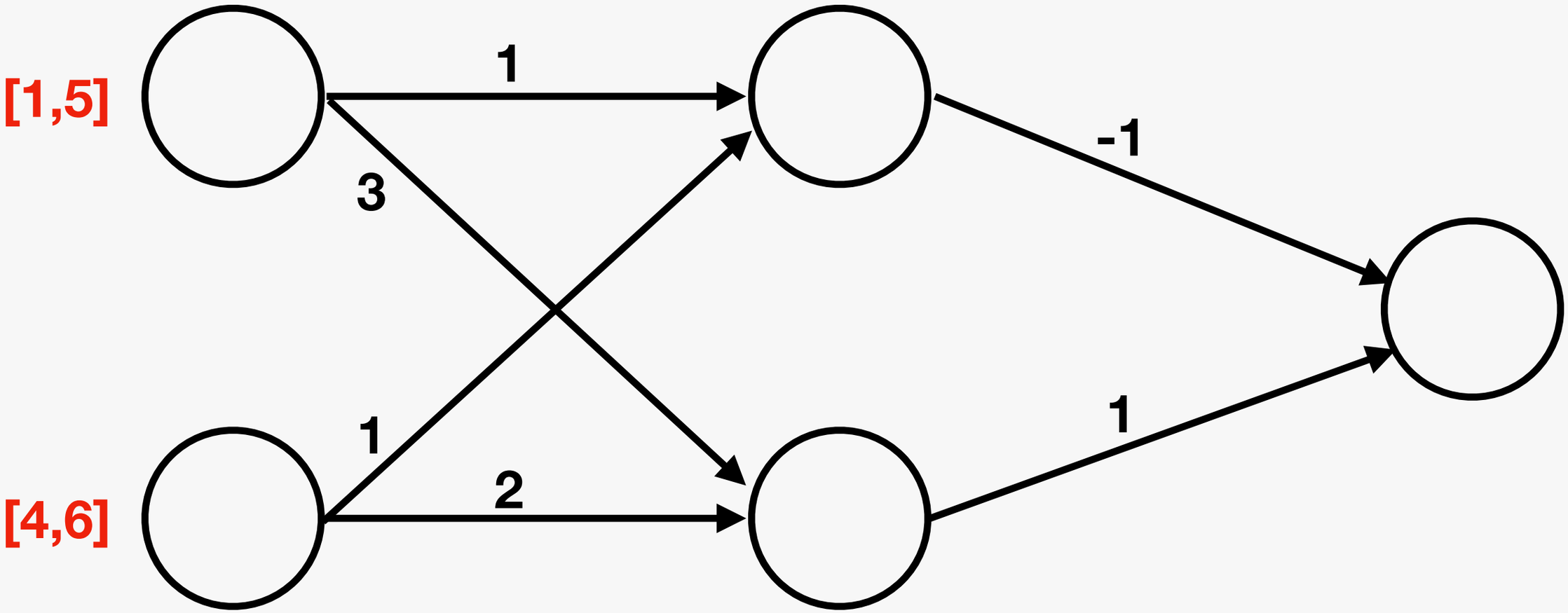
구간 분석 예시

결과값이 20 이하일까?



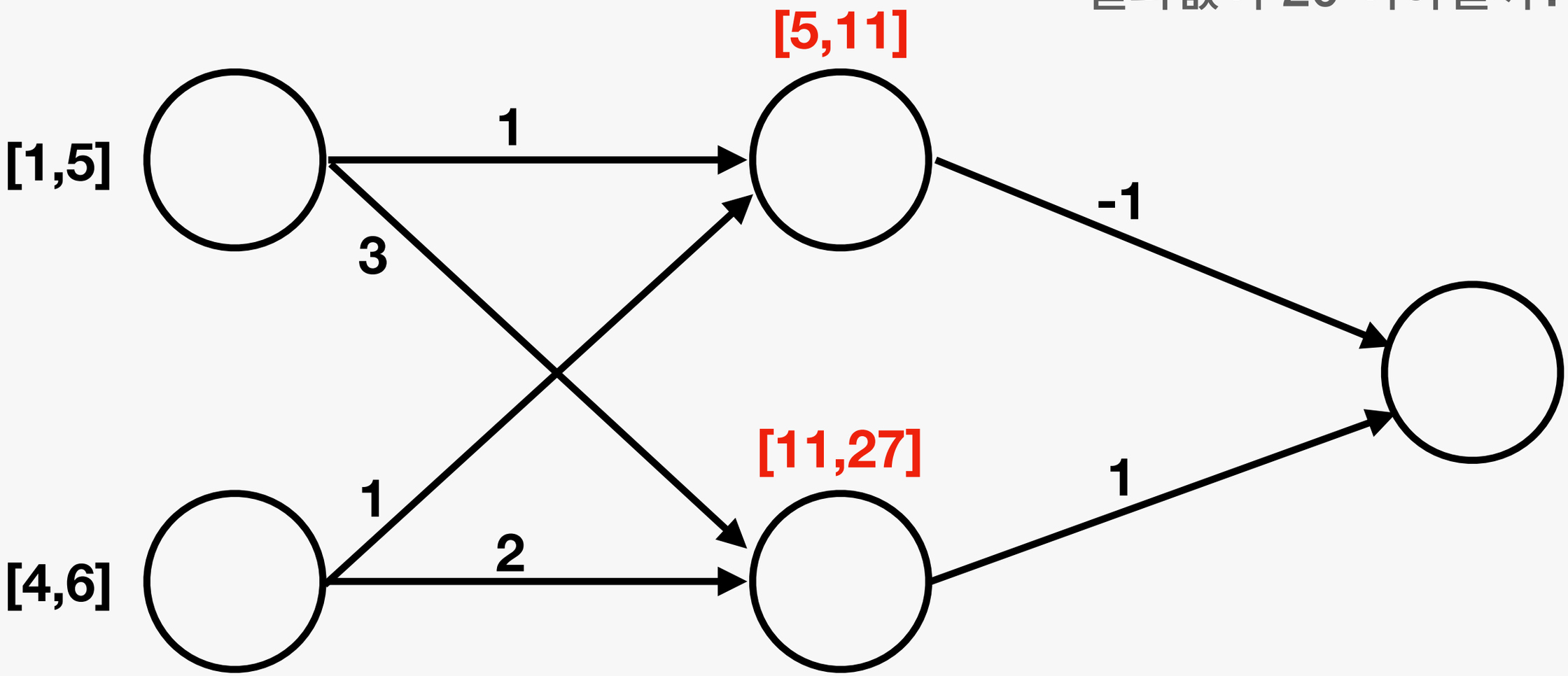
구간 분석 예시

결과값이 20 이하일까?



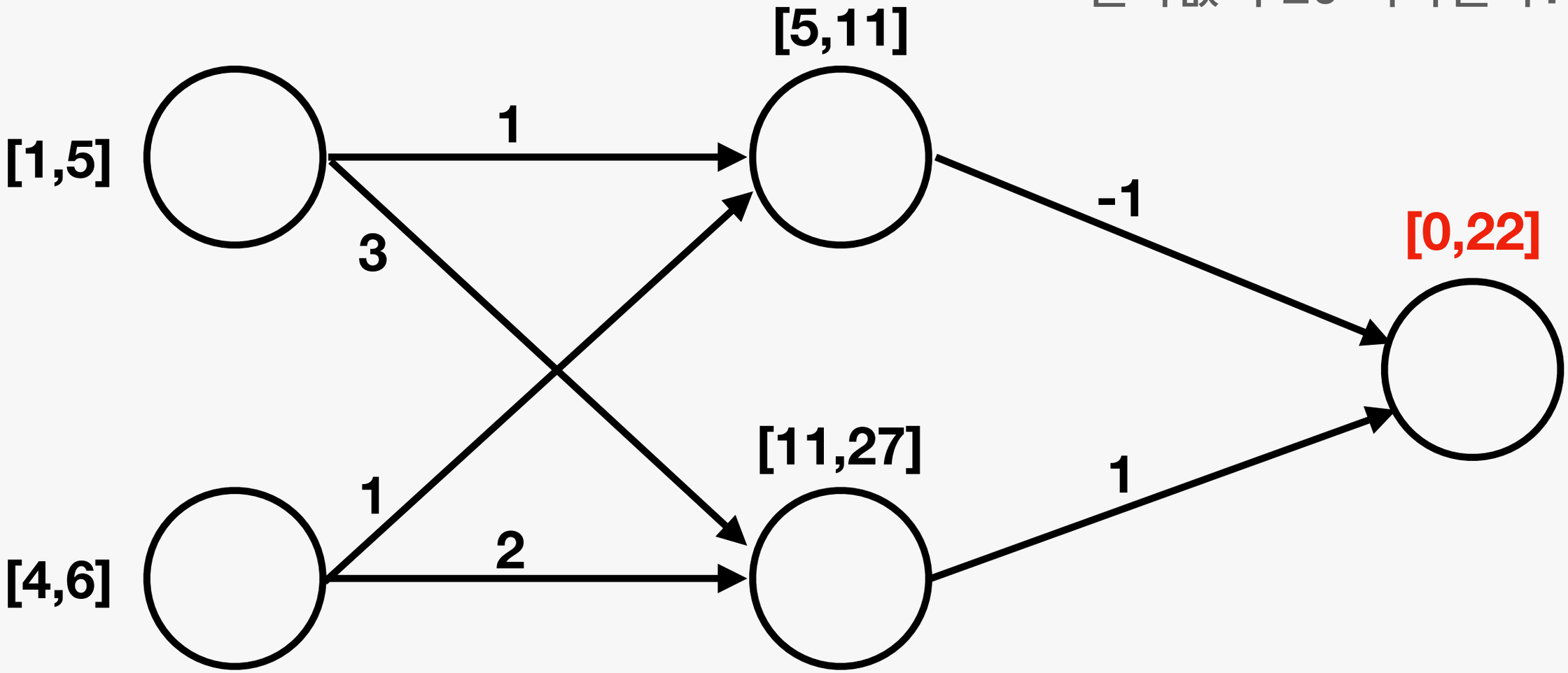
구간 분석 예시

결과값이 20 이하일까?



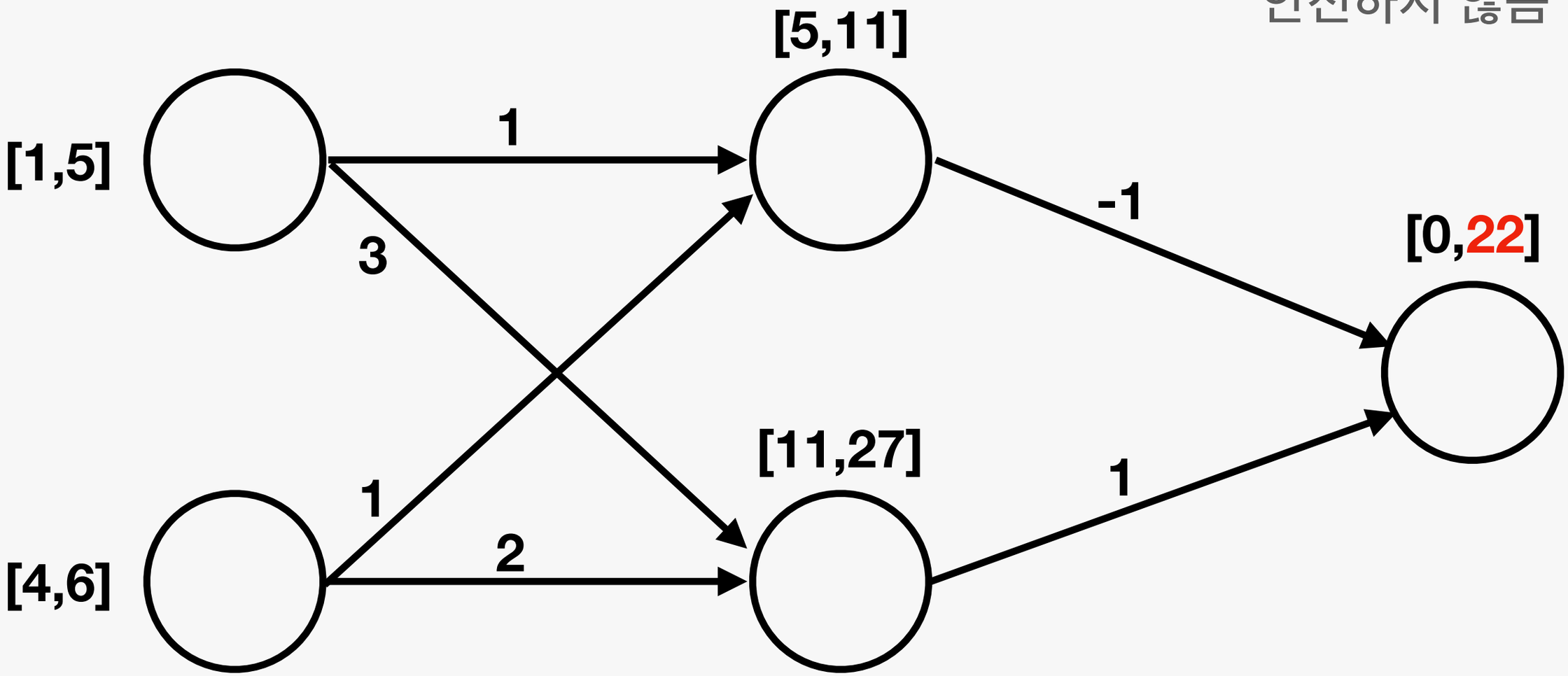
구간 분석 예시

결과값이 20 이하일까?



구간 분석 예시

신경망이
안전하지 않음

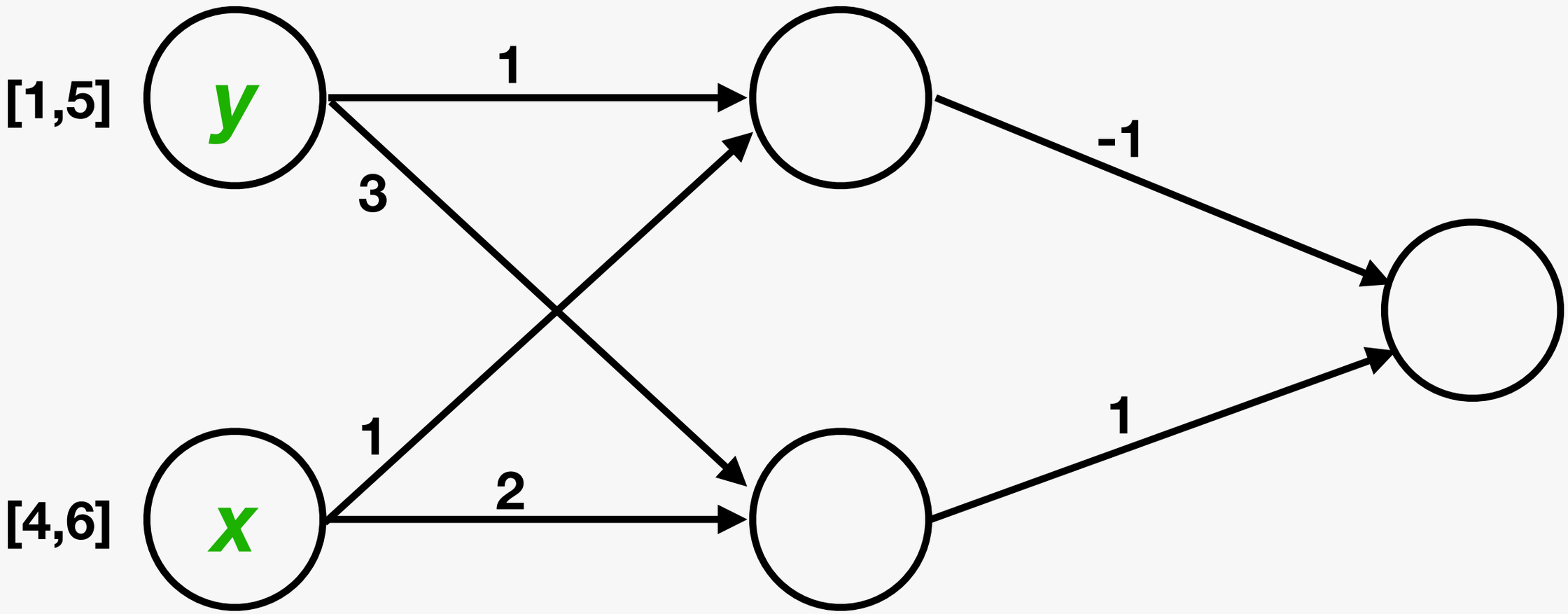


구간 분석 예시

- 계산 결과가 너무 느슨함
  - 0과 22는 실제로 나타날 수 없음!
- 느슨한 결과는 검증의 실패로 나타날 수 있음
  - 다양한 최적화 방법 필요

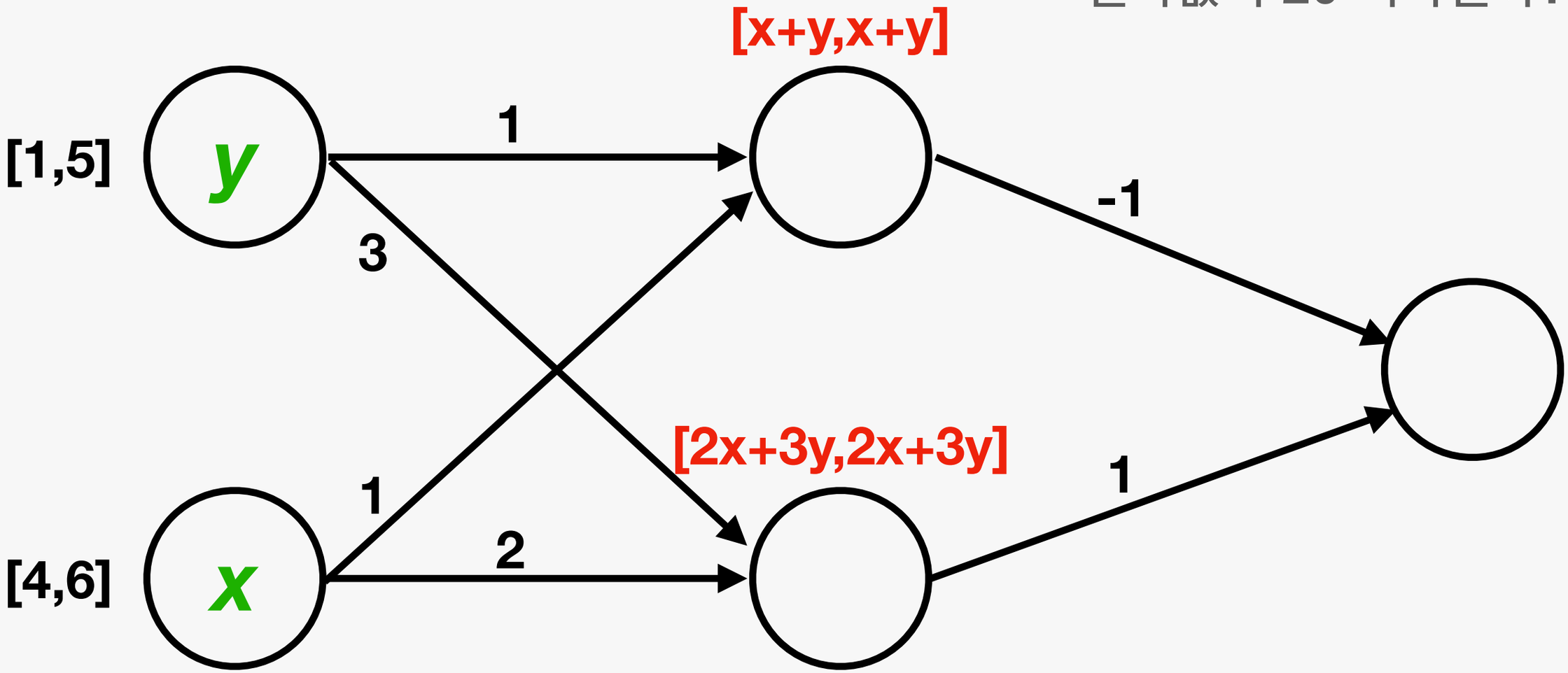
구간 분석 최적화 - (1)

결과값이 20 이하일까?



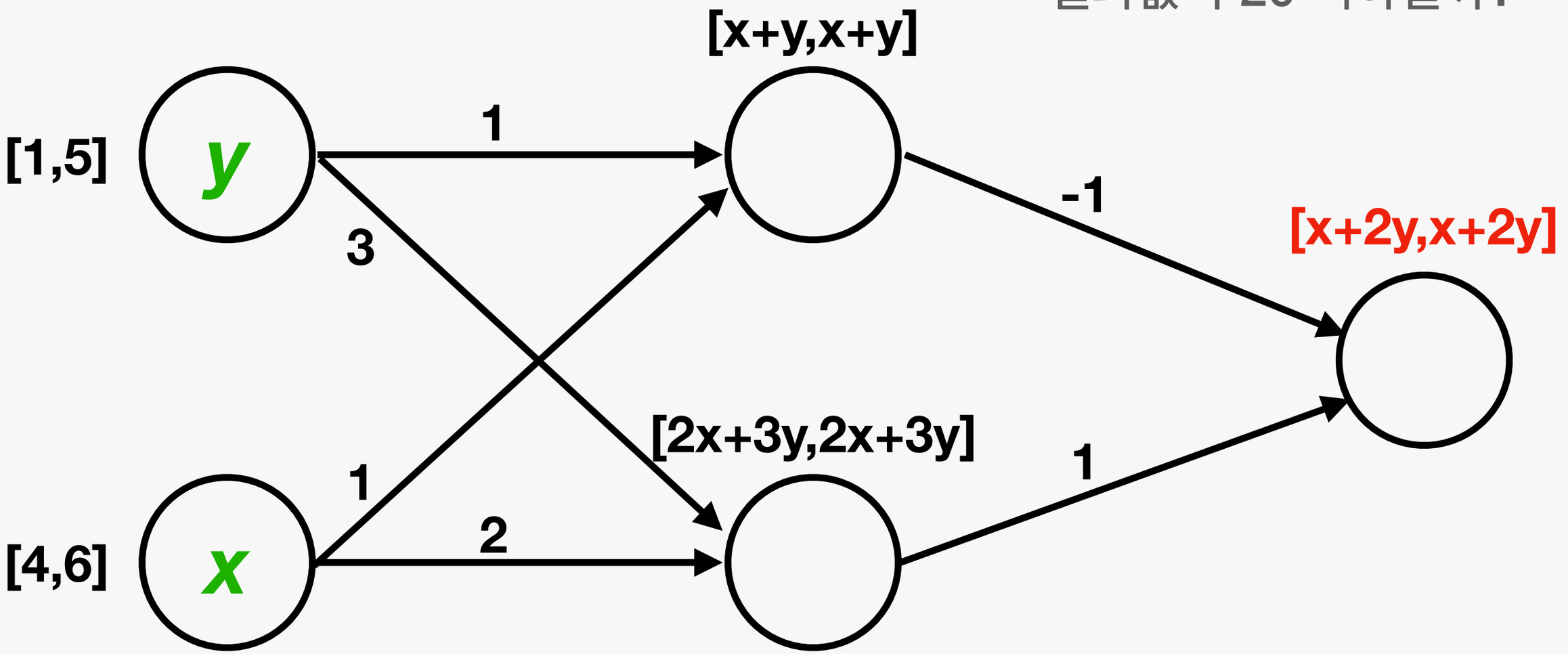
구간 분석 최적화 - (1)

결과값이 20 이하일까?



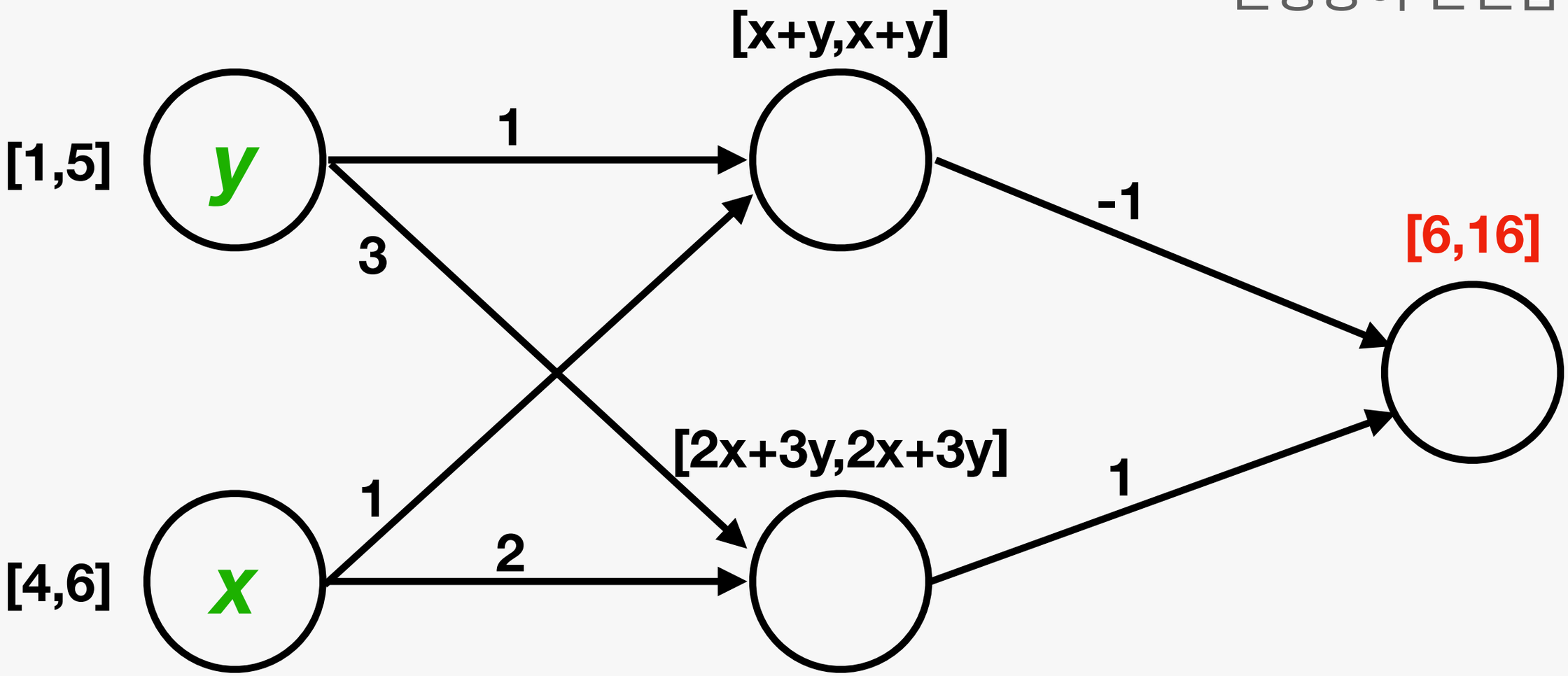
구간 분석 최적화 - (1)

결과값이 20 이하일까?



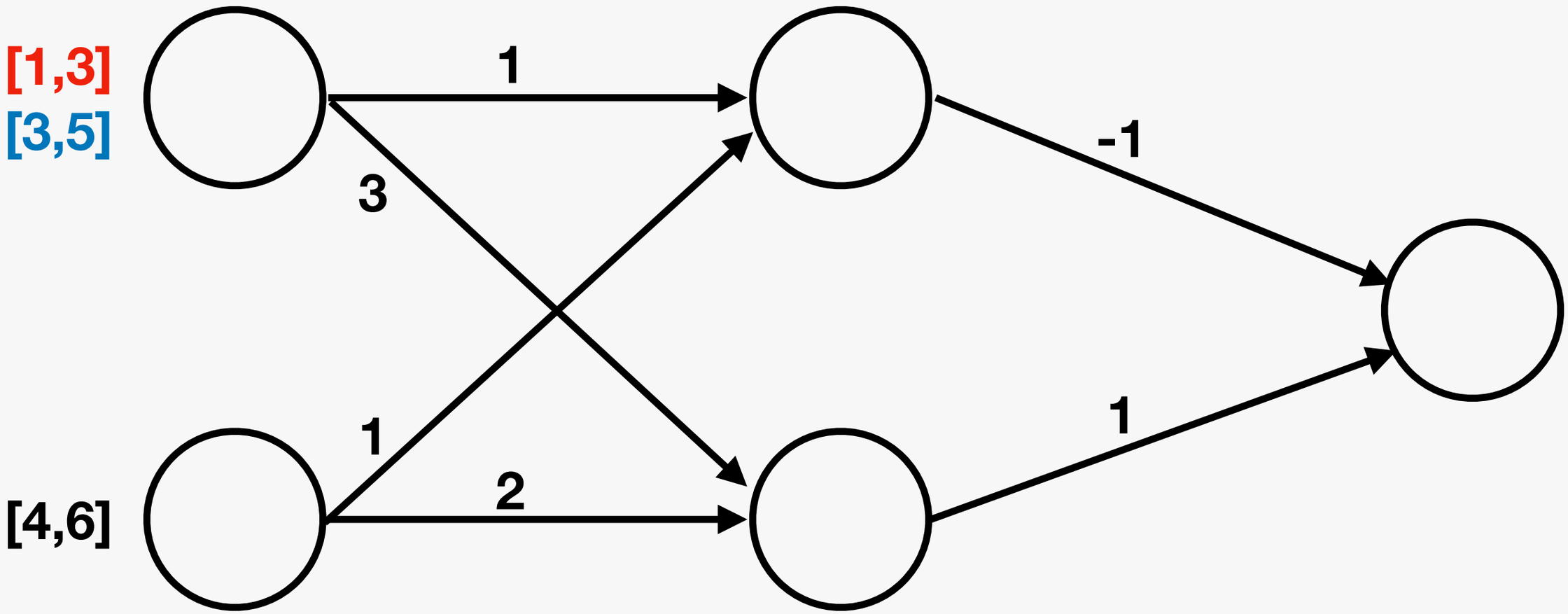
구간 분석 최적화 - (1)

신경망이 안전함

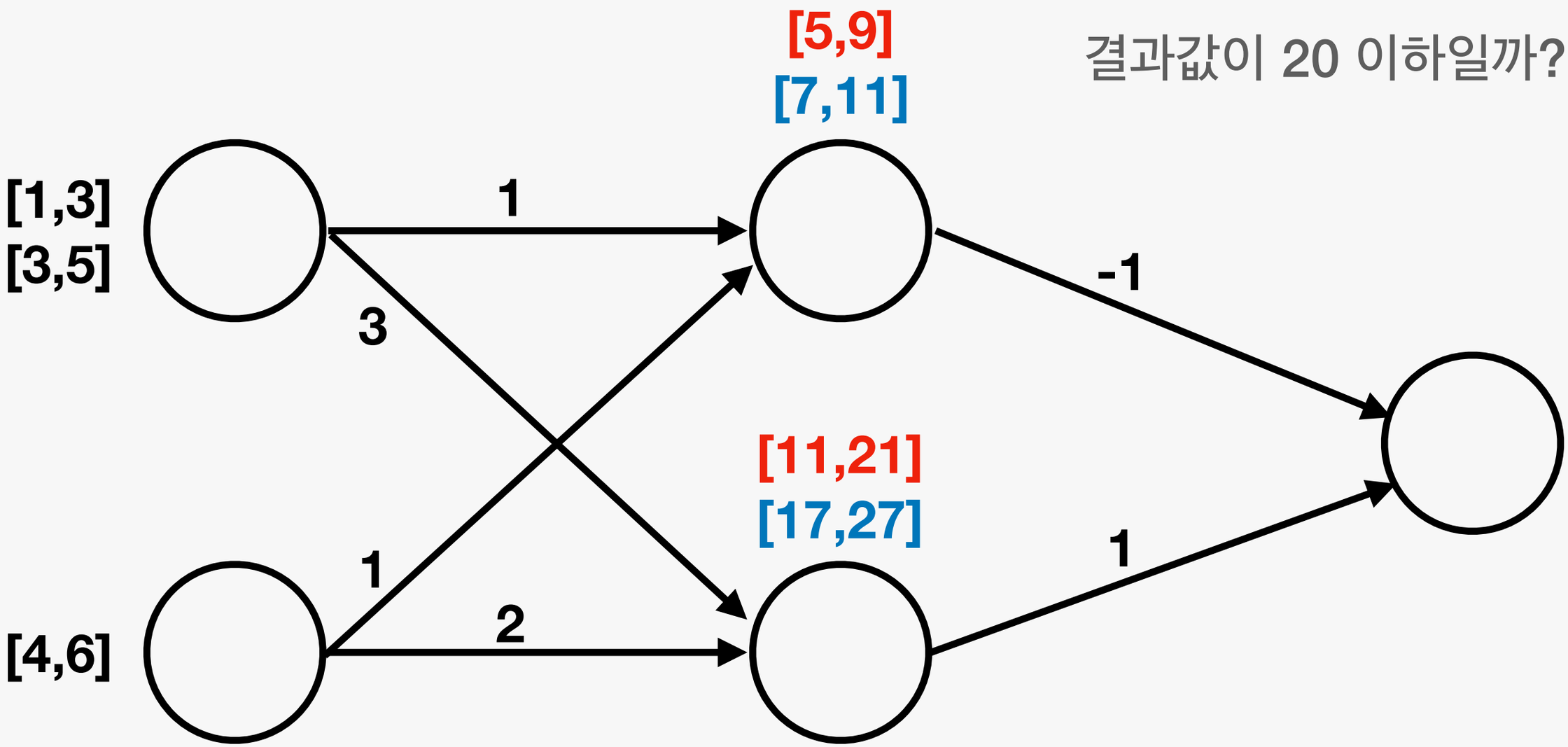


구간 분석 최적화 - (2)

결과값이 20 이하일까?



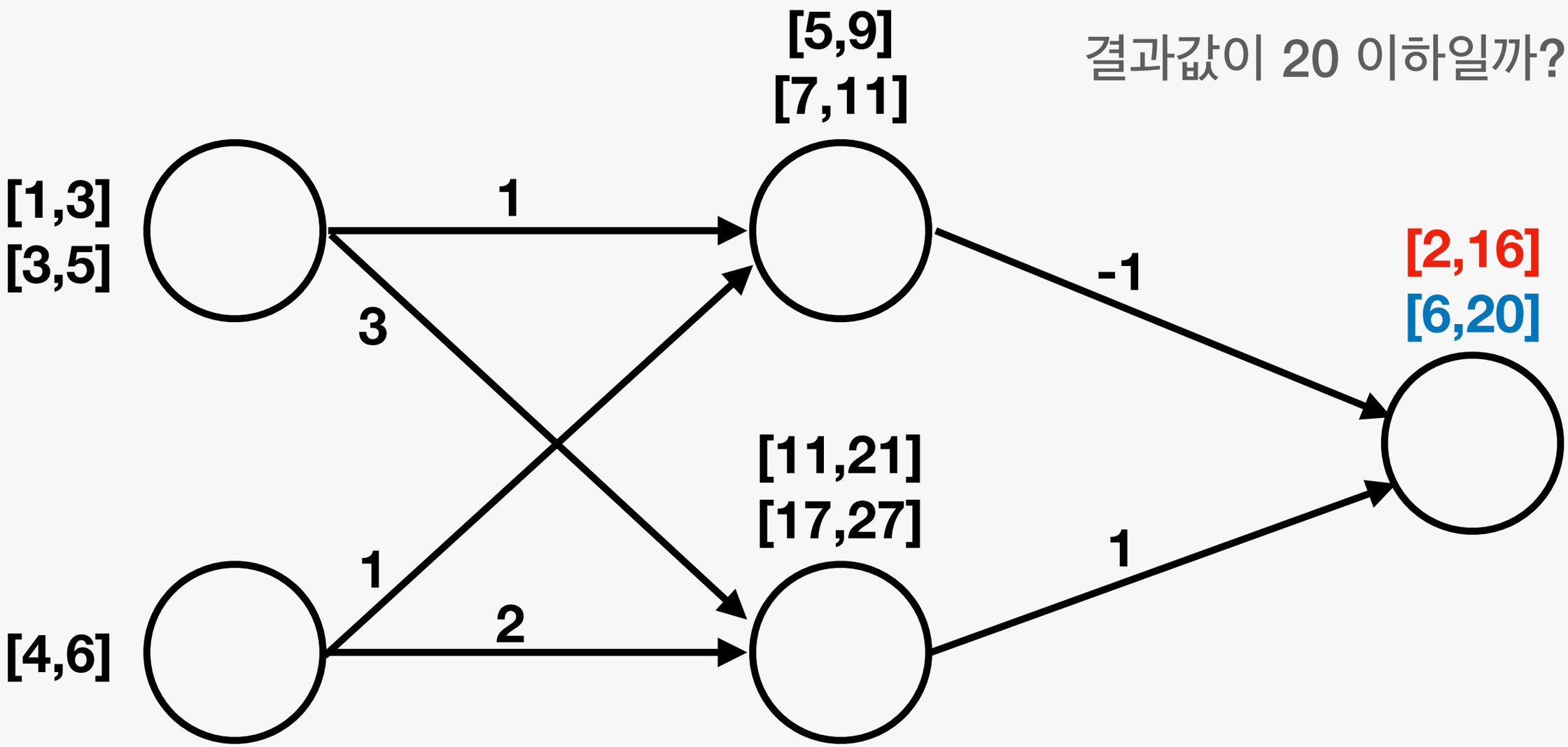
구간 분석 최적화 - (2)



결과값이 20 이하일까?

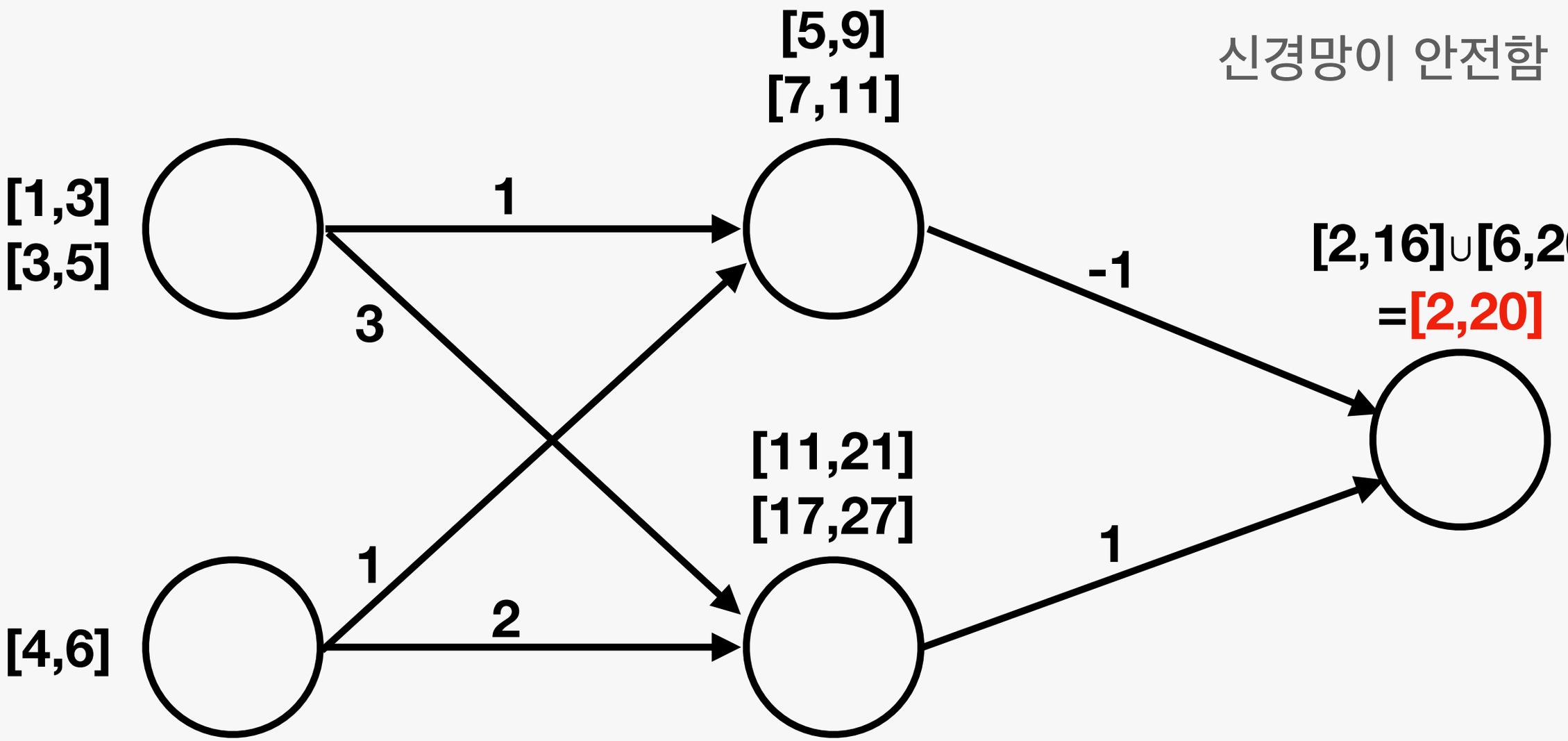
구간 분석 최적화 - (2)

결과값이 20 이하일까?

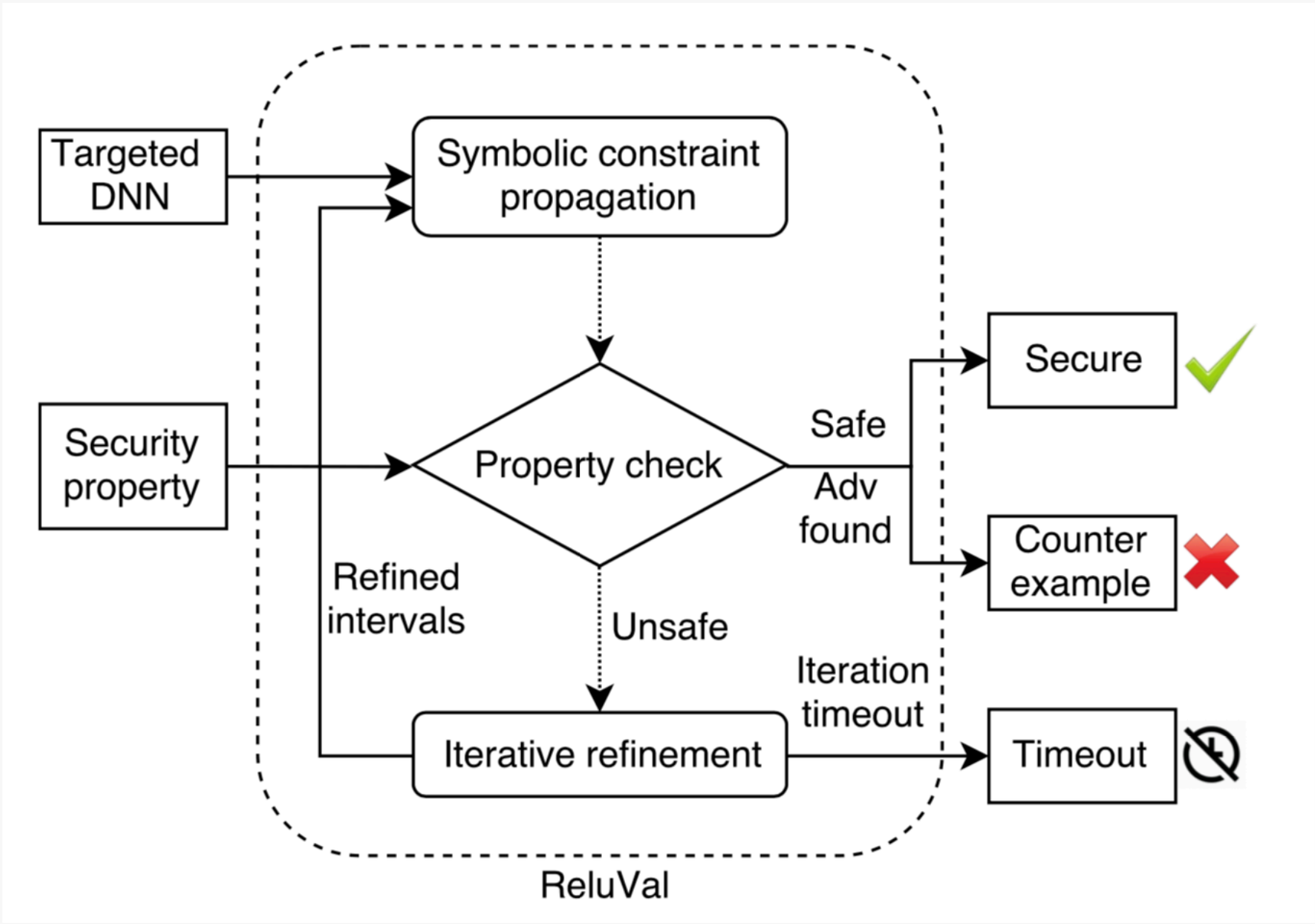


구간 분석 최적화 - (2)

신경망이 안전함



ReLUVal



[Shiqi Wang *et al.*, 2018]

실험 결과

- ACAS Xu. 모델에 대해 실험
  - 항공기의 충돌 방지 시스템
  - 45개의 신경망으로 구성
  - 5개의 입출력, 각 50개의 뉴런을 가진 6개의 은닉층
  - 15개의 안전성 검사

실험 결과

- ReLUPlex*에 비해 평균적으로 200x 이상 빠름
- 적대적인 예시를 찾는 것도 효과적
    - Carlini Wagner 공격보다 더 많은 예시를 찾음

*[G. Katz *et al*., 2017]

더 나아갈 점

- 구간 계산의 정확도, 속도 향상
- 병렬 처리 개선
- 다양한 활성 함수(sigmoid, tanh…)
- 찾아낸 반례들을 신경망 학습에 이용하기

*[G. Katz *et al*., 2017]