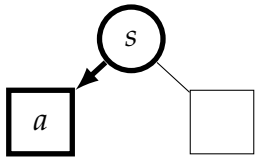


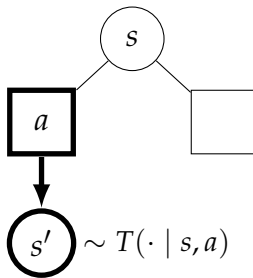
SELECTION



select action that maximizes

$$Q(s, a) + c \sqrt{\frac{\log N(s)}{N(s, a)}}$$

EXPANSION



SIMULATION/ROLLOUT BACKPROPAGATION

