

동적계획법과 강화학습 강의 노트(11월 10일 강의)

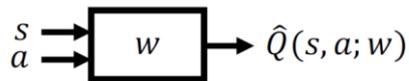
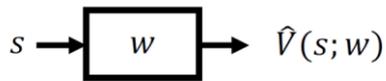
작성자 : 서재완(2021321341)

작성일시 : 2022.11.16

Value Function Approximation

1. Value Function Approximation

State 수가 너무 많게 되면 테이블 형태로 모든 state을 가질 수 없기 때문에 weight 만을 가져 storage 문제를 해결하여 value function을 정확한 값을 찾는 것이 아닌 추정하는 것



2. How to approximate

실제 $V(S)$ 와 예측 값인 $v(S, w)$ 의 mean-squared-error를 통해 w 값을 찾음

$$J(\mathbf{w}) = \mathbb{E}_{\pi} [(v_{\pi}(S) - \hat{v}(S, \mathbf{w}))^2]$$

찾는 방식은 경사하강법을 통해 찾아감

$$\Delta \mathbf{w} = \alpha (v_{\pi}(S) - \hat{v}(S, \mathbf{w})) \nabla_{\mathbf{w}} \hat{v}(S, \mathbf{w})$$

3. Linear VFA

선형 Value Function 추정 방식은 선형 방정식을 이용해 찾아감 마찬가지로 mse 를 통해 경사하강법을 사용하여 w 값을 찾아감

$$\hat{v}(S, \mathbf{w}) = \mathbf{x}(S)^{\top} \mathbf{w} = \sum_{j=1}^n x_j(S) w_j$$

$$\nabla_{\mathbf{w}} \hat{v}(S, \mathbf{w}) = \mathbf{x}(S)$$

$$\Delta \mathbf{w} = \alpha (v_{\pi}(S) - \hat{v}(S, \mathbf{w})) \mathbf{x}(S)$$

4. Incremental Prediction Algorithm

실제 v 값을 안다고 가정했지만 강화학습에서는 리워드만 존재할 뿐 실제 y 값을 알 수 없음. 이에 따라 대체하여 문제를 해결

MC

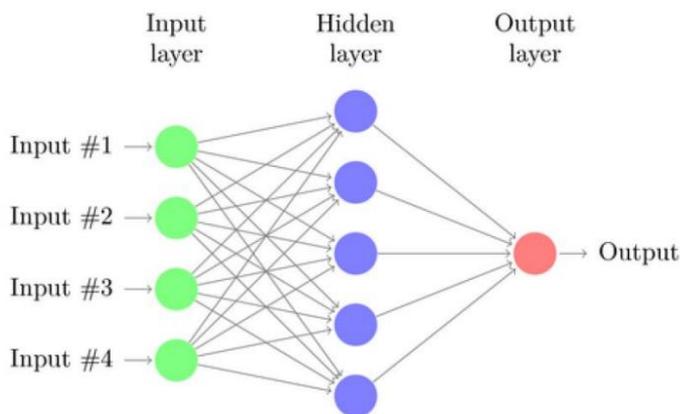
$$\Delta \mathbf{w} = \alpha (G_t - \hat{v}(S_t, \mathbf{w})) \nabla_{\mathbf{w}} \hat{v}(S_t, \mathbf{w})$$

TD

$$\Delta \mathbf{w} = \alpha (R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}) - \hat{v}(S_t, \mathbf{w})) \nabla_{\mathbf{w}} \hat{v}(S_t, \mathbf{w})$$

5. DNN

선형이 아닌 경우 비선형 함수를 추정하기 위해 사용하는 대표적인 모델 중 하나로 인경신경망 모델



여러개의 노드를 통해 여러 함수로 구성되어 있으며 loss 함수를 최소화 하기위해 경사하강법을 통해 기울기를 계산하여 움직일 방향을 정하고 오차역전파를 통해 가중치를 학습해 나간다.

6. CNN

이미지에 적용하는 대표적인 뉴럴네트워크 방법론 중 하나

이미지에 기존의 DNN방식을 적용하면 굉장히 많은 파라미터를 가지지만 CNN을 통해 커널 함수를 사용하여 가중치를 공유하며 학습하고 지역 정보를 잘 학습이 가능함