Title ~~Q-learning~~ ~~Actor-Critic~~ ✓ ~~Policy Gradient Method~~ Date / Time 24/7/11 Page No.

score $\in (-\infty, 10]$

$\|$

W−L

$$\text{state} = \{\overset{I}{\text{inertia}}, \overset{D}{\text{damping}}, \text{ROM}\}$$

~~chromosome~~ = [P_{1+}, P_{1-}, P_{2+}, P_{2-}, P_{3+}, P_{3-}]

~~separate set~~

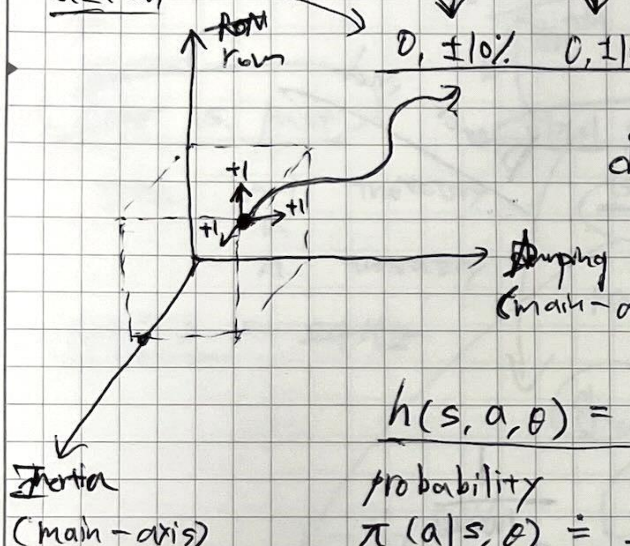
~~separate preferences~~

\+

~~softmax~~

↓

action → 0, ±10% 0, ±10% 0, ±10%



input

3 input dim. → 6 output dim.

$$I\ddot{\theta} + D\dot{\theta} = \tau$$

discrete action space

선호 h → softmax

action

$i + 10\%$ $+0.01$

$i - 10\%$ -0.01

$d + 10\%$

$d - 10\%$ 6

$r + 10\%$

$r - 10\%$

state

$$\begin{pmatrix} i \\ d \\ r \end{pmatrix}$$ 3

$$h(s, a, \theta) = \theta^T x(s, a)$$

probability

$$\pi(a|s, \theta) \doteq \frac{e^{h(s,b,\theta)}}{\sum_b e^{h(s,b,\theta)}}$$

$\hat{v}(s, w)$

$$\frac{\frac{df(x)}{dx}}{x} \qquad \frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)}$$

Lot의 simulation 100.에

\+

noise

$$h(s, a, \theta) = \begin{bmatrix} \omega_1 s_1 + \omega_2 s_2 + \omega_3 s_3 \\ \\ \\ \end{bmatrix} = 6 \begin{bmatrix} i+ \\ i- \\ d+ \\ d- \\ r+ \\ r- \end{bmatrix}$$

pref.

Title Date / Time Page No.

~~Weight (θ the~~

~~θ = weigh~~

linear

$$S = \begin{bmatrix} I \\ D \\ R \end{bmatrix} \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix} \qquad I, D, R \in \mathbb{R}$$

weight $\quad W^T$ $\qquad$ preference $\qquad$ Policy (prob)

$$\begin{bmatrix} w_{I1} & w_{D1} & w_{R1} \\ w_{I2} & w_{D2} & w_{R2} \\ & \vdots & \\ w_{I6} & w_{D6} & w_{R6} \end{bmatrix} \begin{bmatrix} I \\ D \\ R \end{bmatrix} = \begin{bmatrix} \circ \\ \circ \\ \circ \\ \circ \\ \circ \\ \circ \end{bmatrix} \xrightarrow{\text{soft max.}} \begin{bmatrix} \circ \\ \circ \\ \circ \\ \circ \\ \circ \\ \circ \end{bmatrix}$$

$$W = \begin{bmatrix} w_{I1} & \cdots & w_{I6} \\ w_{D1} & \cdots & w_{D6} \\ w_{R1} & \cdots & w_{R6} \end{bmatrix}$$

θ

$$\begin{bmatrix} \theta_{11} & \theta_{21} & \theta_{31} \\ & \vdots & \\ \theta_{16} & \theta_{26} & \theta_{36} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} \theta_{11} s_1 + \theta_{21} s_2 + \theta_{31} s_3 \\ \vdots \\ \theta_{16} s_1 + \theta_{26} s_2 + \theta_{36} s_3 \end{bmatrix} = \begin{bmatrix} h_1 \\ \vdots \\ h_6 \end{bmatrix}$$

$$\pi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_6 \end{bmatrix} = \begin{bmatrix} \cancel{e^{h_1} / (e^{h_1} + \cdots + e^{h_6})} \\ \vdots \\ \cancel{e^{h_6} / (e^{h_1} + \cdots + e^{h_6})} \end{bmatrix}$$

soft max 이용

Title Date / Time Page No.

~~weber-fraction scheduling~~ ✗ α scheduling

시간상수 $e^{-\alpha t}$

action (a)

$s_1 + 10\%$
$+ 0\%$ softmax.
$- 10\%$

$s_2 + 10\%$
$+ 0\%$ softmax.
$- 10\%$

$s_3 + 10\%$
$+ 0\%$ softmax.
$- 10\%$

Weber's ~~rule~~ law

weight perception

$\sim 2\%$

~~weber fraction~~

state

s_1

s_2

s_3

$$S = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

$$\theta^T S = \begin{bmatrix} \theta_{11} s_1 + \theta_{21} s_2 + \theta_{31} s_3 \\ \vdots \\ \theta_{19} s_1 + \theta_{29} s_2 + \theta_{39} s_3 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_9 \end{bmatrix} = h$$

$$\theta = \begin{bmatrix} \theta_{11} & \cdots & \theta_{19} \\ \theta_{21} & \cdots & \theta_{29} \\ \theta_{31} & — & \theta_{39} \end{bmatrix}$$

$$\pi = \begin{bmatrix} \dfrac{e^{h_1}}{e^{\cancel{h_1+h_2+h_3}}} \\ + \\ \vdots \end{bmatrix}$$

$0 < md < 1$

$$\begin{bmatrix} e^{h_1}/(e^{h_1+h_2+h_3}) \\ e^{h_2}/(e^{h_1+h_2+h_3}) \\ e^{h_3}/(e^{h_1+h_2+h_3}) \end{bmatrix}$$

0.1	$md < 0.1$
0.2	$0.1 < md < 0.3$
0.7	$0.3 < md$ ($0.3 + 0.7$)

$$f(x, y, z) = \frac{e^z}{e^x + e^y + e^z}$$

$$\frac{\partial f}{\partial x} =$$

$$\nabla \pi = \begin{bmatrix} \dfrac{\partial \pi}{\partial \theta_{11}} & \cdots & \dfrac{\partial \pi}{\partial \theta_{19}} \\ \vdots & & \vdots \\ \dfrac{\partial \pi}{\partial \theta_{31}} & \cdots & \dfrac{\partial \pi}{\partial \theta_{39}} \end{bmatrix}$$

$\pi_{11} = \dfrac{\partial \pi}{\partial \theta_{11}}$

$\cancel{\pi_{nm} = \dfrac{\partial \pi}{\partial \theta_{n1}}}$

$$\pi_{ij} = \frac{\partial}{\partial \theta_{ij}} \pi$$

Title Date / Time Page No.

$$\frac{\partial}{\partial\theta_{11}}\left(\frac{e^{h_1}}{e^{h_1}+e^{h_2}+e^{h_3}}\right) = \pi_1$$

$$= \cancel{\frac{\partial h_1}{\partial\theta_{11}}}\frac{\partial\pi}{\partial h_1} + \frac{\partial h_2}{\partial\theta_{11}}\frac{\partial\pi}{\partial h_2} + \frac{\partial h_3}{\partial\theta_{11}}\frac{\partial\pi}{\partial h_3}$$

$$= S_1$$

$$h_1 = \theta_{11}S_1 + \theta_{21}S_2 + \theta_{31}S_3$$

$$h_2 = \theta_{12}S_1 + \theta_{22}S_2 + \theta_{32}S_3$$

$$h_3 = \theta_{13}S_1 + \theta_{23}S_3 + \theta_{33}S_3$$

$$\frac{\partial\pi}{\partial h_1} = \frac{e^{h_1}(e^{h_2}+e^{h_3})}{\{e^{h_1}+e^{h_2}+e^{h_3}\}^2} = \cancel{\frac{e^{h_1}(e^{h_2}+e^{h_3})}{\pi^2}}$$

$$= \pi^2\,\frac{e^{h_2}+e^{h_3}}{e^{h_1}}$$