April 13 (Tue)

Reward Function Example: Ref.

a github 105d ; b ML-Robot-Arm

Example: Reward Function, pp.1, Section.

$$S \times A = (s_1 a_1, s_1 a_2, \cdots, s_n a_n)$$

$$S = \{s_1, s_2\}$$

$$A = \{a_1, a_1\}$$

$$S \times A = \{s_1, s_2\} \times \{a_1, a_2\}$$

$$= \{s_1 a_1, s_1 a_2, s_2 a_1, s_2 a_2\}$$

Reward $r : S \times A \rightarrow \underline{R}$

whole collection of all reward

$$R = \{r_1, r_2, r_3, r_4\}$$

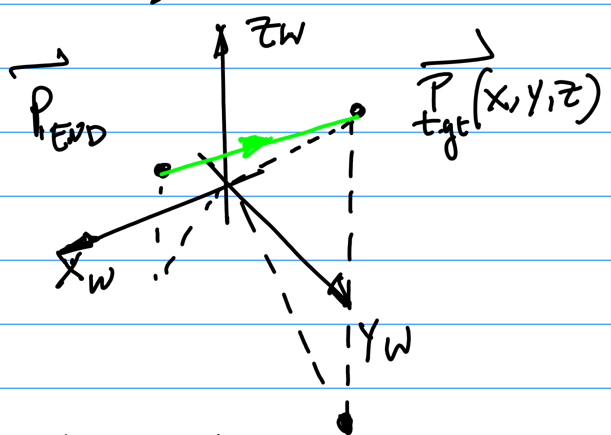
$s_1 a_1$ $s_1 a_2$ ... $s_2 a_2$ ... (1)

Denote a reward at time t as

$$R(s_t, a_t) = R(\text{Hit the Ground}, a_t) = -1 \quad \cdots (2)$$

$$R(s_t, a_t) = R(\text{Reachs the target}, a_t) = 1 \quad \cdots (3)$$

$\vec{P}(x,y,z)$ in X_w-Y_w-Z_w World Coordinate:



$\vec{P}_{tgt}$, $\vec{P}_{END}$ ← End Effector

Move $\vec{P}_{END}$ to $\vec{P}_{tgt}$ By $\underline{N}$ steps

$$(s_t, a_t) \text{ for } t = 1, 2, \dots, N \quad \cdots (3b)$$

if $\vec{P}_{END} = \vec{P}_{tgt}$ then End the Episod.

$$\vec{P}_{END} \simeq \vec{P}_{tgt}, \quad \vec{P}_{END} - \vec{P}_{tgt} \simeq 0 \quad \cdots (4)$$

$$\| \vec{P}_{END} - \vec{P}_{tgt} \| \ll \epsilon \quad \cdots (5)$$

$$\vec{P}_{END}(x,y,z) = \vec{P}_{END}(x(t), y(t), z(t))$$

for $t=1$,

$$\vec{P}_{END}(x(1), y(1), z(1))$$

for $t=2$

$$\vec{P}_{END}(x(2), y(2), z(2))$$

for $t=i$

$$\vec{P}_{END}(x(i), y(i), z(i))$$

$\vdots$

$$\vec{P}_{END}(x(N), y(N), z(N))$$

for $t-1$ to t, the distance (movement of $\vec{P}_{END}$)

$$\vec{P}_{END}(x(t), y(t), z(t)) - \vec{P}_{END}(x(t-1), y(t-1), z(t-1)) \quad \cdots (6)$$

Suppose

$$\vec{P}_{END}(x(0), y(0), z(0)) = (1, 1, 2)$$

$$\vec{P}_{END}(x(1), y(1), z(1)) = (0.5, -1, 2)$$

Find the distance

$$\vec{P}_{END}(x(1), y(1), z(1)) - \vec{P}_{END}(x(0), y(0), z(0)) = (1-0.5, -1-1, 2-2) = (0.5, -2, 0)$$

$$\|\vec{P}_{END}(x(1), y(1), z(1)) - \vec{P}_{END}(x(0), y(0), z(0))\|_2$$

$$= \sqrt{(x(1)-x(0))^2 + (y(1)-y(0))^2 + (z(1)-z(0))^2}$$

$$= \sqrt{0.5^2 + (-2)^2 + 0^2} = \sqrt{4.25}$$ Distance to the target

$d = \|\vec{P}_{END}(t-1) - \vec{P}_{tgt}\|_2$, @ $t-1$

$\downarrow$ $d(t-1)$

$d(t) = \|\vec{P}_{END}(t) - \vec{P}_{tgt}\|_2$, @ t

$$R(s_t, a_t) = \frac{d(\vec{P}_{END}(t-1) - \vec{P}_{tgt}) - d(\vec{P}_{END}(t) - \vec{P}_{tgt})}{d(\vec{P}_{END}(0) - \vec{P}_{tgt})} \quad \cdots (9*)$$

Difference of the distance

Assume $d(\vec{P}_{END}(t-1) - \vec{P}_{tgt}) = \sqrt{6.25}$ Positive Reward.

$d(\vec{P}_{END}(t) - \vec{P}_{tgt}) = \sqrt{4.25}$

$$\tilde{R}(s_t, a_t) = \sqrt{6.25} - \sqrt{4.25}$$

$\downarrow$ Normalization

$$R(s_t, a_t) = \tilde{R} / d(\vec{P}_{END}^{(0)} - \vec{P}_{tgt})$$

For Negative Reward, from Eqn (9*) holds good as well, Since Eqn (9*) will give negative value.

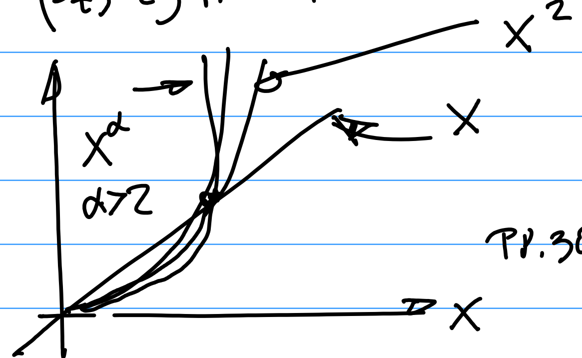
$$R(s_t, a_t) = \frac{d(\vec{P}_{END}(t-1) - \vec{P}_{tgt}) - d(\vec{P}_{END}(t) - \vec{P}_{tgt})}{d(\vec{P}_{END}(0) - \vec{P}_{tgt})}$$

$$-1 \le R(s_t, a_t) \le 1$$

$$-1 \le R^2(s_t, a_t) \le 1 \Rightarrow 0 \le R^2(s_t, a_t) \le 1$$

$$\| R^{\alpha}(s_t, a_t) \| \le 1$$

Let $R = X$



Pg.30

(1) Code | Source Walk-Through (2) Dynamic Programming

Non-Linear Accelerated Reward Function with $\alpha > 2$

Question: How About $e^{\pm R}$ or $\log_a R$