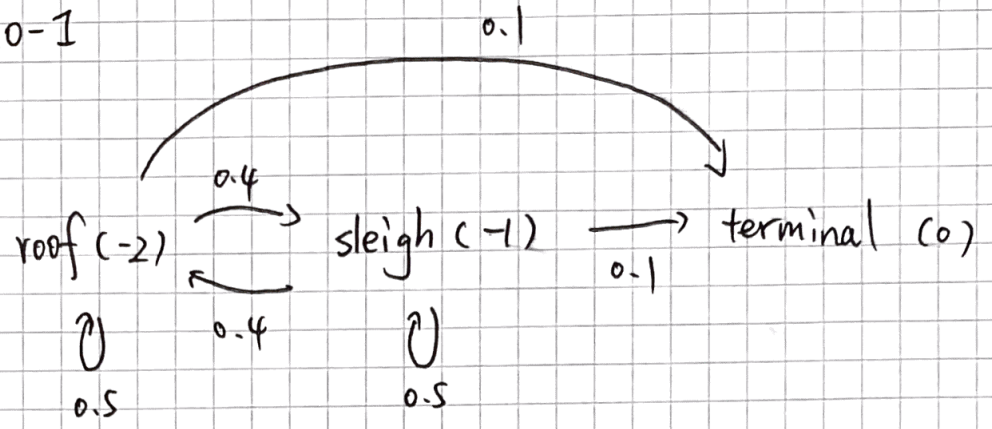
Ex 10-1



(a) Transition probability:

	sleigh	roof	terminal
sleigh	0.5	0.4	0.1
roof	0.4	0.5	0.1
terminal	0	0	0

$$P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}$$

(b) Bellman equation:

$$U(s) = R(s) + \gamma \sum_{s'} p(s'|s)\, U(s') \quad , \forall s \in S.$$

$$\Rightarrow \quad U = R + \gamma P U \Rightarrow (I - \gamma P) U = R.$$

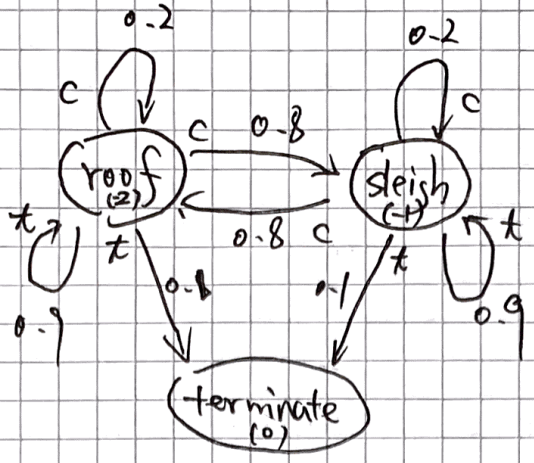
$$\Rightarrow \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0 & 0 & 0 \end{pmatrix} \right] \begin{bmatrix} u_s \\ u_r \\ u_t \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}$$

$$\Rightarrow \quad u_s = 30 - u_r = \frac{130}{9}$$

$$\begin{cases} u_r = -\frac{140}{9} \\ u_t = 0 \end{cases}$$

Ex 10-2:

(a)



$p(X_r|X_r,c) = 0.2$ $\quad p(X_r|X_r,t) = 0.9$

$p(X_r|X_s,c) = 0.8$ $\quad p(X_r|X_s,t) = 0.1$

$p(X_s|X_s,c) = 0.2$ $\quad p(X_s|X_s,t) = 0.9$

$p(X_s|X_r,c) = 0.8$ $\quad p(X_s|X_r,c) = 0.1$

(b) sleigh cost lower, jump to sleigh then terminate.

(c) Initial policy π_0 $\pi_0(X_r) = t$, $\pi_0(X_s) = t$.

Policy evaluation: $U^{\pi_t}(x) = U^t(x) = R(x) + \gamma \sum_{s'} p(x'|x, \pi_t(x)) U^t$

① policy evaluation:

$$\begin{cases} u_s = -1 + 0.1u_t + 0.9u_s \\ u_r = -2 + 0.1u_t + 0.9u_r \\ u_t = 0 \end{cases} \Rightarrow \begin{cases} u_s = -10 \\ u_r = -20 \\ u_t = 0 \end{cases}$$

② poliy improvement: $\pi(s) = \arg\max_a \sum_{s' \in S} p(s'|s,a) U^*(s')$

$T(x_i, a) = \sum_{s' \in S} p(s'|s,a) U^*(s')$

$T(X_s, c) = 0.2 \times (-10) + 0.8 \times (-20) = -18$

$T(X_r, c) = 0.2 \times (-20) + 0.8 \times (-10) = -12$

$T(X_s, t) = 0.1 \times (0) + 0.9 \times (-10) = -9$

$T(X_r, t) = 0.1 \times (0) + 0.9 \times (-20) = -18$

$T(X_s, c) < T(X_s, t) \Rightarrow \pi_1(X_s) = t$.

$T(X_r, c) > T(X_r, t) \Rightarrow \pi_1(X_r) = c$.

π_1 is changed from π_0.

③ policy evaluation:

$$\begin{cases} u_s = -1 + 0.1u_t + 0.9u_s \\ u_r = -2 + 0.2u_r + 0.8u_s \\ u_t = 0 \end{cases} \Rightarrow \begin{cases} u_s = -10 \\ u_r = -12.5 \\ u_t = 0 \end{cases}$$

④ policy improvement.

$T(X_s, c) = 0.8 \times (-12.5) + 0.2 \times (-10) = -12$

$T(X_s, t) = 0.9 \times (-10) + 0.1 \times 0 = -9$

$T(X_r, c) = 0.8 \times (-10) + 0.2 \times (-12.5) = -10.5$

$T(X_r, t) = 0.9 \times (-12.5) + 0.1 \times 0 = -11.25$.

$T(X_s, c) < T(X_s, t) \Rightarrow \pi_2(X_s) = t$.

$T(X_r, c) > T(X_r, t) \Rightarrow \pi_2(X_r) = c$.

π_2 is not change compare to π_1. terminate.

Big belly $\Leftarrow$ All presents from sleigh. instead of claiming.

(d) Initial policy $\pi_0(X_r) = c$ $\pi_0(X_s) = c$.

$$\begin{cases} u_s = -1 + 0.2u_s + 0.8u_r \\ u_r = -2 + 0.8u_s + 0.2u_r \\ u_t = 0 \end{cases} \Rightarrow -\infty$$