Assignment 4

> NAME: 董骏博
>
> SID: 12432995

Task 1

1. Initialize 2 center points by random, here I choose `MLP` and `CNN` as the initial center points
2. Iteration
   - Compute the distance between every single point with the 2 center points
   - Assign data points to the category of the nearest center point
   - repeat compute the new center point of every category
   - stop iteration if center points keep same
3. get 2 center points and 2 categories

Task 2

For original MDP:

$$V(s) = \max_{a \in A} \sum_{s' \in S} P(s'|s,a)[R(s) + \gamma V(s')]$$

and it can be write with optimal policy π^* as below

$$V(s) = \sum_{s' \in S} P(s'|s,\pi^*(s))[R(s) + \gamma V(s')]$$

For the modified MDS with new reward function:

$$\begin{aligned} V'(s) &= \max_{a \in A} \sum_{s' \in S} P(s'|s,a)[\alpha R(s) + \beta + \gamma V(s')] \\ &= \alpha \max_{a \in A} \sum_{s' \in S} P(s'|s,a)[R(s) + \frac{\beta}{\alpha} + \gamma \frac{V'(s')}{\alpha}] \end{aligned}$$

and it can be write with new optimal policy π^** as below

$$\begin{aligned} V'(s) &= \sum_{s' \in S} P(s'|s,\pi^{**}(s))[(\alpha R(s) + \beta) + \gamma V'(s')] \\ &= \alpha \sum_{s' \in S} P(s'|s,\pi^{**}(s))[R(s) + \frac{\beta}{\alpha} + \gamma \frac{V'(s')}{\alpha}] \end{aligned}$$

Because $\alpha > 0$, maximizing the value function $V(s)$ will also maximize the policy π^*. The policy maximized will not be infected when reward was changed by the above formulates. So, the modified MDP will has the same optimal policy as the original MDP.

Task 3

(1)

For `Operational` state:

$$\begin{aligned} V_O &= 0.9(1+\gamma V_O) + 0.1(0+\gamma V_F) \\ &= \frac{0.9+0.1\gamma V_F}{1-0.9\gamma} \end{aligned}$$

For `Faulty` state:

$$\begin{aligned} V_F &= 0.1(-10+\gamma V_F) + 0.9(0+\gamma V_O) \\ &= \frac{0.9\gamma V_O - 1}{1-0.1\gamma} \end{aligned}$$

(2)

For `Operational` state:

$$V^*(n) = \max\left\{0.9(1+\gamma V^*(n)) + 0.1(0+\gamma V^*(a)), 1 \times (0+\gamma V^*(n))\right\}$$

For `Faulty` state:

$$V^*(a) = \max\left\{0.9(0+\gamma V^*(n)) + 0.1(-10+\gamma V^*(a)), 0\right\}$$

And, we can get:

For `Operational` state:

$$\begin{cases} \text{take 'I' action} & 0.9 + 0.9\gamma V^*(n) + 0.1\gamma V^*(a) > \gamma V^*(n) \\ \text{take 'D' action} & \text{otherwise} \end{cases}$$

For `Faulty` state:

$$\begin{cases} \text{take 'R' action} & 0.9\gamma V^*(n) - 1 + 0.1\gamma V^*(a) > 0 \\ \text{no action} & \text{otherwise} \end{cases}$$