HW5

學號:107062631　　姓名:方晟軒

1. TODO

基本上照著下面演算法實作 TODO 的部分即可完成

function REINFORCE
Initialise θ arbitrarily
for each episode $\{s_1, a_1, r_2, ..., s_{T-1}, a_{T-1}, r_T\} \sim \pi_\theta$ **do**
for $t = 1$ to $T - 1$ **do**
$\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(s_t, a_t) v_t$
end for
end for
return θ
end function

```
        loss = 0
        for i in range(steps):
            # TODO:
            # Take out state, action, reward from SAR_list
            # Compute loss
            sar = SAR_list[i]
            cur_state = sar.state
            cur_action = sar.action
            cur_reward = rewards[i]

            probs = policy_net(cur_state)           # <= hint: feed something into policy network
            m = Categorical(probs)
            loss += -m.log_prob( cur_action ) * cur_reward
            # END TODO
```

把助教所提供的部分，將問號依序填入，就將作業 5 完成了。