```
% NPC（学習プレイヤー）の行動選択と，学習する際の対戦相手となる人工知能プレイヤーのプログラム

function [action,reward,state3,fin]=action_train(policy,step,state3)
  reward = 0;

  % 学習プレイヤー
  % 最初のステップでは1マス目を選択
  if(step == 1)
    a=1;
  else
    % 政策policyに従いランダムに行動を選択
    while(1)
      random = rand;

      cprob = 0;
      for a=1:9
        cprob = cprob + policy(a);
        if(random < cprob)
          break;
        end
      end

      % 既にマスが埋まっていないかどうかを確認
      if state3(a)==0
        break;
      end
    end
  end

  action = a;
  state3(a) = 2;
  fin = check(state3);
  if(fin == 2)
    reward = 10;
    return;
  elseif(fin == 3)
    reward = 0;
    return;
  end

  % 人工知能プレイヤー
  reach = 0;
  pos = [1 2 3; 4 5 6; 7 8 9; 1 4 7; 2 5 8; 3 6 9; 1 5 9; 3 5 7];
  for i=1:max(size(pos))
    val = sum(state3(pos(i,:)));
    num = size(find(state3(pos(i,:))==0),2);
    if(val==2 & num==1)
      a = pos(i,state3(pos(i,:))==0);
      reach = 1;
      break;
    end
  end

  if(reach==0)
    while(1)
```

```
    a = floor(rand*9)+1;

    if state3(a)==0
      break;
    end
  end
end

state3(a) = 1;

fin = check(state3);
if(fin == 1)
  reward = -10;
  return;
elseif(fin==3)
  reward = 0;
  return;
end
```