

```

% LEARN_CW_Script – Performs on-policy sarsa and Q-learning for the
% windy grid world example.
%
% Written by:
% --
% John L. Weatherwax          2007-12-03
%
% email: wax@alum.mit.edu
%
% Please send comments and especially bug reports to the
% above email address.
%
%-----

close all;

save_figs = false;

alpha = 1e-1;

sideII = 4; sideJJ = 12;

% create the "cliff":
CF = ones(sideII,sideJJ); CF(sideII,2:(sideJJ-1)) = 0;

% the beginning and terminal states (in matrix notation):
s_start = [ 4, 1 ];
s_end = [ 4, 12 ];

MAX_N_EPISODES=30;
MAX_N_EPISODES=500; % 500 episodes done in the book
%MAX_N_EPISODES=1e4;
MAX_N_EPISODES=1e5;
%MAX_N_EPISODES=1e6;
%MAX_N_EPISODES=10e6;

[Q_sarsa,Q_qlearn,rpt_sarsa,rpt_qlearn,n_sarsa,n_qlearn] =
learn_cw(alpha,CF,s_start,s_end,MAX_N_EPISODES);

% extract the greedy policy and state value function from both:
%
pol_pi_sarsa = zeros(sideII,sideJJ); V_sarsa = zeros(sideII,sideJJ); n_g_sarsa =
zeros(sideII,sideJJ);
pol_pi_qlearn = zeros(sideII,sideJJ); V_qlearn = zeros(sideII,sideJJ); n_g_qlearn =
zeros(sideII,sideJJ);
for ii=1:sideII,
    for jj=1:sideJJ,
        sti = sub2ind( [sideII,sideJJ], ii, jj );
        [V_sarsa(ii,jj),pol_pi_sarsa(ii,jj)] = max( Q_sarsa(sti,:) );
        n_g_sarsa(ii,jj) = n_sarsa(sti,pol_pi_sarsa(ii,jj));
        [V_qlearn(ii,jj),pol_pi_qlearn(ii,jj)] = max( Q_qlearn(sti,:) );
        n_g_qlearn(ii,jj) = n_qlearn(sti,pol_pi_qlearn(ii,jj));
    end
end

% sarsa:
%
plot_cw_policy(pol_pi_sarsa,CF,s_start,s_end);
title( 'sarsa policy' );

```

```
fn = sprintf('cw_sarsa_policy_nE_%d',MAX_N_EPISODES); if( save_figs ) saveas( gcf, fn,
'png' ); end

figure; imagesc( V_sarsa ); colormap(flipud(jet)); colorbar;
title( 'sarsa state value function' );
fn = sprintf('cw_sarsa_state_value_fn_nE_%d',MAX_N_EPISODES); if( save_figs ) saveas(
gcf, fn, 'png' ); end

figure; imagesc( n_g_sarsa ); colorbar;
title( 'SARSA: number of greedy samples' );

% Plot the reward per episode as in the book:
rpe_sarsa = cumsum(rpt_sarsa)./cumsum(1:length(rpt_sarsa));
ph=figure; ph_sarsa = plot( rpe_sarsa, '-x' ); axis([0, 1000, -5 0]); grid on; hold on;

% Qlearn:
%
plot_cw_policy(pol_pi_qlearn,CF,s_start,s_end);
title( 'qlearn policy' );
fn = sprintf('cw_qlearn_policy_nE_%d',MAX_N_EPISODES); if( save_figs ) saveas( gcf, fn,
'png' ); end

figure; imagesc( V_qlearn ); colormap(flipud(jet)); colorbar;
title( 'qlearn state value function' );
fn = sprintf('cw_qlearn_state_value_fn_nE_%d',MAX_N_EPISODES); if( save_figs ) saveas(
gcf, fn, 'png' ); end

figure; imagesc( n_g_qlearn ); colorbar;
title( 'QLEARN: number of greedy samples' );

rpe_qlearn = cumsum(rpt_qlearn)./cumsum(1:length(rpt_qlearn));
figure(ph); ph_qlearn = plot( rpe_qlearn, '-og' );
xlabel('total episodes'); ylabel('Reward per episode'); title('Average reward per
episode');
legend([ph_sarsa,ph_qlearn],{'SARSA','QLEARN'}, 'location', 'southeast');
fn = sprintf('cw_avg_reward_per_episode_nE_%d',MAX_N_EPISODES);
if( save_figs ) saveas( gcf, fn, 'png' ); end
```