[Addendum] Q-Learning

Jae Yun JUN KIM*

March 25, 2021

Algorithm 1: Q-Learning algorithm

initialize $Q(s,a) = 0 \qquad \forall s, \quad \forall a$;
while *not convergenced* **do**
  choose randomly s_t (for $t = 0$);
  while *stop condition is not satisfied* **do**
    $m = \text{rand}()$;
    if $(m < \epsilon)$ **then**
      choose randomly a_t;
    else
      $a_t = \arg\max_a Q(s_t, a)$;
    obtain (s_{t+1}, r_t) from (s_t, a_t);
    $Q(s_t, a_t) \longleftarrow (1-\alpha) \cdot Q(s_t, a_t) + \alpha\,[r_t + \gamma \max_a Q(s_{t+1}, a)]$;
    $s_t \longleftarrow s_{t+1}$
return Q;

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr