Les ensembles

- θ : parameters
- $\mathcal{A}$: action set
- $\mathcal{S}$: State set
- $\mathcal{V}$: The value-function
- $\mathcal{Q}$: the Q-value
- $\Phi(s,a)$: accessible states
- $\mathcal{T}(s'|s,a)$: Transition function
- $\Gamma(s) = \{a \in \mathcal{A} | \exists s' \in \mathcal{S} / \mathcal{T}(s'|s,a) \neq 0\}$
- $\pi_\theta(s)$: policy function
- $r(s,a,s')$: instant reward
- $R(s,a) = \sum_{s' \in \Phi(s,a)} \mathcal{T}(s'|s,a) * r(s,a,s')$

Formules

- $(e_i)_{i \in [\![1;N]\!]} / \forall i \in [\![1;N]\!], e_i = (S_t^i, A_t^i, R_t^i = r(S_{t-1}^i, A_{t-1}^i, S_t^i))_{t \in [\![1,L_i]\!]}$
- $C_s = \{(i,t) / S_t^i = s\}$
- $U_{s,a} = \{(i,t) \in C_s / A_t^i = a\}$
- $K_{s,a,s'} = \{(i,t) \in U_{s,a} / S_{t+1}^i = s'\}$
- $G_t^i = \sum_{k=0}^{L_i - t} \gamma^k R_{t+k}^i$: Gain function
- $\mathcal{T}(s'|s,a) \approx \frac{|K_{s,a,s'}|}{|U_{s,a}|}$
- Bellman's equation for the Valuation function

$$\mathcal{V}(s) = \frac{1}{|C_s|} \sum_{(i,t) \in C_s} G_t^i \tag{1}$$

$$\mathcal{V}(s) = \frac{1}{|C_s|} \sum_{(i,t) \in C_s} R_t^i + \gamma G_{t+1}^i \tag{2}$$

$$\mathcal{V}(s) = \frac{1}{|C_s|} \sum_{(i,t) \in C_s} R_t^i + \gamma V(S_{t+1}^i) \tag{3}$$

$$\mathcal{V}(s) = \sum_{a \in \Gamma(s)} \frac{|U_{s,a}|}{\sum_{a' \in \Gamma(s)} |U_{s,a}|} * \mathcal{Q}(s,a) \tag{4}$$

$$\mathcal{V}(s) = \sum_{a \in \Gamma(s)} \pi(a|s) \mathcal{Q}(s,a) \tag{5}$$

- Bellmans's equation for the Q value function

$$\mathcal{Q}(s,a) = \frac{1}{|U_{s,a}|} \sum_{(i,t) \in U_{s,a}} G_t^i \tag{6}$$

$$\mathcal{Q}(s,a) = \frac{1}{|U_{s,a}|} \sum_{(i,t) \in U_{s,a}} R_t^i + \gamma G_{t+1}^i \tag{7}$$

$$\mathcal{Q}(s,a) = \frac{1}{|U_{s,a}|} \sum_{(i,t) \in U_{s,a}} R_t^i + \gamma V(S_{t+1}^i) \tag{8}$$

$$\mathcal{Q}(s,a) = \frac{1}{|U_{s,a}|} \sum_{(i,t) \in U_{s,a}} R_t^i + \gamma \sum_{a' \in \Gamma(S_{t+1}^i)} \pi(a'|S_{t+1}^i) \mathcal{Q}(S_{t+1}^i, a') \tag{9}$$

$$\mathcal{J}(\theta) = \sum_{s \in \mathcal{S}} d^\pi(s) \mathcal{V}^\pi(s)$$

$$\nabla_\theta \mathcal{J}(\theta) = \nabla_\theta \sum_{s \in \mathcal{S}} d^\pi(s) \sum_{a \in \mathcal{A}} \mathcal{Q}^\pi(s,a) \pi_\theta(a|s)$$

$$\nabla_\theta \mathcal{J}(\theta) = \propto \sum_{s \in \mathcal{S}} d^\pi(s) \sum_{a \in \mathcal{A}} \mathcal{Q}^\pi(s,a) \nabla_\theta \pi_\theta(a|s)$$