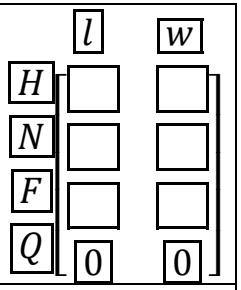
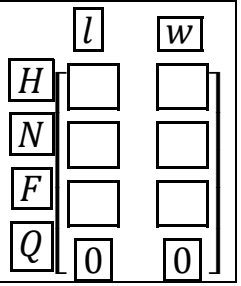
RL Tutorial worksheet

Exercise 1 – Q iteration:

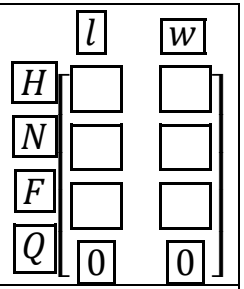
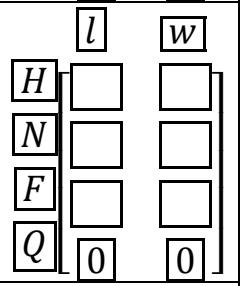
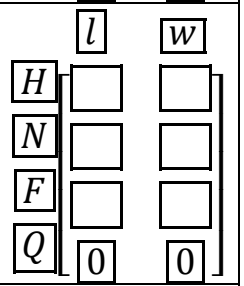
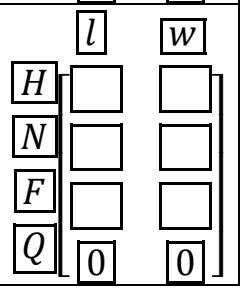
$$\boldsymbol{Q_{new}}(s_t, a_t) \leftarrow \boldsymbol{Q}(s_t, a_t) + \alpha(r_t + \gamma \max_a \boldsymbol{Q}(s_{t+1}, a) - \boldsymbol{Q}(s_t, a_t))$$

$$\boldsymbol{Q_0} = \begin{array}{c} \\ H \\ N \\ F \\ Q \end{array} \begin{array}{c} \begin{array}{cc} l & w \end{array} \\ \begin{bmatrix} 10 & 8 \\ 10 & 8 \\ 10 & 8 \\ 0 & 0 \end{bmatrix} \end{array} \qquad \begin{aligned} \alpha &= 0.8 \\ \gamma &= 0.9 \end{aligned}$$

User 1: $N \xrightarrow{l} F \xrightarrow{l} Q$:

Observation 1: $(s_t, a_t, r_t, s_{t+1}) = (N, l, +1, F)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_1} =$ 
Observation 2: $(s_t, a_t, r_t, s_{t+1}) = (\ldots, \ldots, \ldots, \ldots)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_2} =$ 

User 2: $H \xrightarrow{l} N \xrightarrow{l} N \xrightarrow{l} F \xrightarrow{w} N$:

Observation 3: $(s_t, a_t, r_t, s_{t+1}) = (\ldots, \ldots, \ldots, \ldots)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_3} =$ 
Observation 4: $(s_t, a_t, r_t, s_{t+1}) = (\ldots, \ldots, \ldots, \ldots)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_4} =$ 
Observation 5: $(s_t, a_t, r_t, s_{t+1}) = (\ldots, \ldots, \ldots, \ldots)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_5} =$ 
Observation 6: $(s_t, a_t, r_t, s_{t+1}) = (\ldots, \ldots, \ldots, \ldots)$ $\boldsymbol{Q_{new}}(\ldots, \ldots) \leftarrow \boldsymbol{Q}(\ldots, \ldots) + \alpha(\ldots + \gamma \max_a \boldsymbol{Q}(\ldots, a) - \boldsymbol{Q}(\ldots, \ldots))$	$\boldsymbol{Q_6} =$ 

Exercise 2 – Value Iteration:

$$\boldsymbol{V_{new}}(s) \leftarrow \max_a \left\{ \sum_{s'} P(s'|s,a) \left(r(s,a,s') + \gamma \boldsymbol{V}(s') \right) \right\}, \quad \gamma = 0.9$$

$$\boldsymbol{V_0} = \begin{matrix} H \\ N \\ F \\ Q \end{matrix} \begin{bmatrix} 10 \\ 10 \\ 10 \\ 0 \end{bmatrix}$$

		Next State			
		H	N	F	Q
Current State	H	0.8	0.1	0.0	0.1
	N	0.0	0.5	0.2	0.3
	F	0.0	0.0	0.2	0.8
When user loses: a = l					

		Next State			
		H	N	F	Q
Current State	H	0.9	0.0	0.0	0.1
	N	0.7	0.2	0.0	0.1
	F	0.3	0.4	0.1	0.2
When user wins: a = w					

$\boldsymbol{V_{new}}(F)$

$\leftarrow \max \begin{Bmatrix} \dots\dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + P(Q|F,l) \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \\ \dots\dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + P(Q|F,w) \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \end{Bmatrix}$

$= \dots\dots\dots$

$\boldsymbol{V_{new}}(N)$

$\leftarrow \max \begin{Bmatrix} P(H,N,l) \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \\ P(H,N,w) \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \end{Bmatrix}$

$= \dots\dots\dots$

$\boldsymbol{V_{new}}(H)$

$\leftarrow \max \begin{Bmatrix} \dots\dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \\ \dots\dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + \cdots . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) + . \dots\dots \times (\dots\dots\dots + \gamma \times \dots\dots\dots) \end{Bmatrix}$

$= \dots\dots\dots$