Tutorial 3.2 (a)

CZ3005

Recall that the set of all the states $\mathcal{S} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3)\}$, the set of actions $\mathcal{A} = \{\mathbf{N}orth, \mathbf{S}outh, \mathbf{W}est, \mathbf{E}ast\}$, the reward discount is $\gamma = 0.9$, $Q(S,A)$, and $V(S)$ are initialized as 0. Suppose we adopt synchronized update (Q and V tables are only updated after each iteration but not within the iteration).

Given the update formula:
$Q(S=s, A=a) \leftarrow \sum_{s' \in \mathcal{S}} P(s'|s,a) \cdot (R(s,a,s') + \gamma V(s'))$,

for the **first iteration**, we update $Q((1,1),N)$ as:

$$
\begin{aligned}
Q\left((1,1),N\right) \leftarrow & \underbrace{P((1,1)|(1,1),N)}_{\text{0.1: left wall collision}} \cdot \left(\underbrace{R((1,1),N,(1,1))}_{\text{0: non-terminal reward is always 0}} + 0.9 \cdot \underbrace{V((1,1))}_{\text{0: as initialized}} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(1,1),N)}_{0.1} \cdot \left(\underbrace{R((1,1),N,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(1,1),N)}_{\text{0: impossible to jump over grid}} \cdot \left(\underbrace{R((1,1),N,(1,3))}_{\text{-5: negative terminal}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(1,1),N)}_{0.8} \cdot \left(\underbrace{R((1,1),N,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& {\color{blue} 0.8 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(1,1),N)}_{0} \cdot \left(\underbrace{R((1,1),N,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(1,1),N)}_{0} \cdot \left(\underbrace{R((1,1),N,(2,3))}_{\text{5: positive terminal}} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned}
\tag{1}
$$

We update $Q((1,1),S)$ as:

$$\begin{aligned}
Q\left((1,1),S\right) \leftarrow\ & \underbrace{P((1,1)|(1,1),S)}_{\text{0.9: left\&btm wall collision}} \cdot \left(\underbrace{R((1,1),S,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0}\right) \\
& 0.9 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,2)|(1,1),S)}_{0.1} \cdot \left(\underbrace{R((1,1),S,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0}\right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,3)|(1,1),S)}_{0} \cdot \left(\underbrace{R((1,1),S,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0}\right) \\
& 0 \times (-5 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,1)|(1,1),S)}_{0} \cdot \left(\underbrace{R((1,1),S,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0}\right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,2)|(1,1),S)}_{0} \cdot \left(\underbrace{R((1,1),S,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0}\right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,3)|(1,1),S)}_{0} \cdot \left(\underbrace{R((1,1),S,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0}\right) \\
& 0 \times (5 + 0.9 \times 0) = 0 \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned} \tag{2}$$

We update $Q((1,1),W)$ as:

$$
\begin{aligned}
Q\left((1,1),W\right) \leftarrow\ & \underbrace{P((1,1)|(1,1),W)}_{\text{0.9: left\&btm wall collision}} \cdot \left(\underbrace{R((1,1),W,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0}\right) \\
& {\color{blue} 0.9 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(1,1),W)}_{0} \cdot \left(\underbrace{R((1,1),W,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0}\right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(1,1),W)}_{0} \cdot \left(\underbrace{R((1,1),W,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0}\right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(1,1),W)}_{0.1} \cdot \left(\underbrace{R((1,1),W,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0}\right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(1,1),W)}_{0} \cdot \left(\underbrace{R((1,1),W,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0}\right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(1,1),W)}_{0} \cdot \left(\underbrace{R((1,1),W,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0}\right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned} \tag{3}
$$

We update $Q((1,1),E)$ as:

$$
\begin{aligned}
Q\left((1,1),E\right) \leftarrow & \underbrace{P((1,1)|(1,1),E)}_{\text{0.1: btm wall collision}} \cdot \left(\underbrace{R((1,1),E,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(1,1),E)}_{0.8} \cdot \left(\underbrace{R((1,1),E,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& {\color{blue} 0.8 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(1,1),E)}_{0} \cdot \left(\underbrace{R((1,1),E,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(1,1),E)}_{0.1} \cdot \left(\underbrace{R((1,1),E,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(1,1),E)}_{0} \cdot \left(\underbrace{R((1,1),E,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(1,1),E)}_{0} \cdot \left(\underbrace{R((1,1),E,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned}
\tag{4}
$$

Tutorial 3.2 (a)

We update $Q((1,2),N)$ as:

$$
\begin{aligned}
Q\left((1,2),N\right) \leftarrow & \underbrace{P((1,1)|(1,2),N)}_{0.1}\cdot\left(\underbrace{R((1,2),N,(1,1))}_{0}+0.9\cdot\underbrace{V((1,1))}_{0}\right)\\
& 0.1\times(0+0.9\times 0)=0\\
& +\underbrace{P((1,2)|(1,2),N)}_{0}\cdot\left(\underbrace{R((1,2),N,(1,2))}_{0}+0.9\cdot\underbrace{V((1,2))}_{0}\right)\\
& 0\times(0+0.9\times 0)=0\\
& +\underbrace{P((1,3)|(1,2),N)}_{0.1}\cdot\left(\underbrace{R((1,2),N,(1,3))}_{-5}+0.9\cdot\underbrace{V((1,3)}_{0}\right)\\
& 0.1\times(-5+0.9\times 0)=-0.5\\
& +\underbrace{P((2,1)|(1,2),N)}_{0}\cdot\left(\underbrace{R((1,2),N,(2,1))}_{0}+0.9\cdot\underbrace{V((2,1))}_{0}\right)\\
& 0\times(0+0.9\times 0)=0\\
& +\underbrace{P((2,2)|(1,2),N)}_{0.8}\cdot\left(\underbrace{R((1,2),N,(2,2))}_{0}+0.9\cdot\underbrace{V((2,2))}_{0}\right)\\
& 0.8\times(0+0.9\times 0)=0\\
& +\underbrace{P((2,3)|(1,2),N)}_{0}\cdot\left(\underbrace{R((1,2),N,(2,3))}_{5}+0.9\cdot\underbrace{V((2,3))}_{0}\right)\\
& 0\times(5+0.9\times 0)=0\\
& =0+0+(-0.5)+0+0+0=-0.5
\end{aligned}
\tag{5}
$$

We update $Q((1,2),S)$ as:

$$\begin{aligned}
Q\left((1,2),S\right) \leftarrow & \underbrace{P((1,1)|(1,2),S)}_{0.1}\cdot\left(\underbrace{R((1,2),S,(1,1))}_{0}+0.9\cdot\underbrace{V((1,1))}_{0}\right)\\
& 0.1\times(0+0.9\times 0)=0\\
& +\underbrace{P((1,2)|(1,2),S)}_{0.8}\cdot\left(\underbrace{R((1,2),S,(1,2))}_{0}+0.9\cdot\underbrace{V((1,2))}_{0}\right)\\
& 0.8\times(0+0.9\times 0)=0\\
& +\underbrace{P((1,3)|(1,2),S)}_{0.1}\cdot\left(\underbrace{R((1,2),S,(1,3))}_{\text{-5}}+0.9\cdot\underbrace{V((1,3)}_{0}\right)\\
& 0.1\times(-5+0.9\times 0)=-0.5\\
& +\underbrace{P((2,1)|(1,2),S)}_{0}\cdot\left(\underbrace{R((1,2),S,(2,1))}_{0}+0.9\cdot\underbrace{V((2,1))}_{0}\right)\\
& 0\times(0+0.9\times 0)=0\\
& +\underbrace{P((2,2)|(1,2),S)}_{0}\cdot\left(\underbrace{R((1,2),S,(2,2))}_{0}+0.9\cdot\underbrace{V((2,2))}_{0}\right)\\
& 0\times(0+0.9\times 0)=0\\
& +\underbrace{P((2,3)|(1,2),S)}_{0}\cdot\left(\underbrace{R((1,2),S,(2,3))}_{5}+0.9\cdot\underbrace{V((2,3))}_{0}\right)\\
& 0\times(5+0.9\times 0)=0\\
& =0+0+(-0.5)+0+0+0=-0.5
\end{aligned}\tag{6}$$

We update $Q((1,2),W)$ as:

$$
\begin{aligned}
Q\left((1,2),W\right) \leftarrow & \underbrace{P((1,1)|(1,2),W)}_{0.8}\cdot\left(\underbrace{R((1,2),W,(1,1))}_{0}+0.9\cdot\underbrace{V((1,1))}_{0}\right) \\
& {\color{blue}0.8\times(0+0.9\times 0)=0} \\
& +\underbrace{P((1,2)|(1,2),W)}_{0.1}\cdot\left(\underbrace{R((1,2),W,(1,2))}_{0}+0.9\cdot\underbrace{V((1,2))}_{0}\right) \\
& {\color{blue}0.1\times(0+0.9\times 0)=0} \\
& +\underbrace{P((1,3)|(1,2),W)}_{0}\cdot\left(\underbrace{R((1,2),W,(1,3))}_{\text{-5}}+0.9\cdot\underbrace{V((1,3)}_{0}\right) \\
& {\color{blue}0\times(-5+0.9\times 0)=0} \\
& +\underbrace{P((2,1)|(1,2),W)}_{0}\cdot\left(\underbrace{R((1,2),W,(2,1))}_{0}+0.9\cdot\underbrace{V((2,1))}_{0}\right) \\
& {\color{blue}0\times(0+0.9\times 0)=0} \\
& +\underbrace{P((2,2)|(1,2),W)}_{0.1}\cdot\left(\underbrace{R((1,2),W,(2,2))}_{0}+0.9\cdot\underbrace{V((2,2))}_{0}\right) \\
& {\color{blue}0.1\times(0+0.9\times 0)=0} \\
& +\underbrace{P((2,3)|(1,2),W)}_{0}\cdot\left(\underbrace{R((1,2),W,(2,3))}_{5}+0.9\cdot\underbrace{V((2,3))}_{0}\right) \\
& {\color{blue}0\times(5+0.9\times 0)=0} \\
& =0+0+0+0+0+0=0
\end{aligned}
\tag{7}
$$

We update $Q((1,2),E)$ as:

$$
\begin{aligned}
Q\left((1,2),E\right) \leftarrow & \underbrace{P((1,1)|(1,2),E)}_{0} \cdot \left(\underbrace{R((1,2),E,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0}\right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(1,2),E)}_{0.1} \cdot \left(\underbrace{R((1,2),E,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0}\right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(1,2),E)}_{0.8} \cdot \left(\underbrace{R((1,2),E,(1,3))}_{-5} + 0.9 \cdot \underbrace{V((1,3)}_{0}\right) \\
& {\color{blue} 0.8 \times (-5 + 0.9 \times 0) = -4} \\
& + \underbrace{P((2,1)|(1,2),E)}_{0} \cdot \left(\underbrace{R((1,2),E,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0}\right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(1,2),E)}_{0.1} \cdot \left(\underbrace{R((1,2),E,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0}\right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(1,2),E)}_{0} \cdot \left(\underbrace{R((1,2),E,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0}\right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + (-4) + 0 + 0 + 0 = -4
\end{aligned}
\tag{8}
$$

We don't need to update $Q((1,3),A)$ and $Q((2,3),A)$, as the values of the termination states are assumed to be always 0, that is, $V((1,3)) = 0$ and $V((2,3)) = 0$. That is to say, $Q((1,3),A)$ and $Q((2,3),A)$ have nothing to do with the value updates. So, we leave them as initialized (=0).

We update $Q((2,1),N)$ as:

$$
\begin{aligned}
Q\left((2,1),N\right) \leftarrow & \underbrace{P((1,1)|(2,1),N)}_{0} \cdot \left(\underbrace{R((2,1),N,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,2)|(2,1),N)}_{0} \cdot \left(\underbrace{R((2,1),N,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,3)|(2,1),N)}_{0} \cdot \left(\underbrace{R((2,1),N,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& 0 \times (-5 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,1)|(2,1),N)}_{0.9} \cdot \left(\underbrace{R((2,1),N,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& 0.9 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,2)|(2,1),N)}_{0.1} \cdot \left(\underbrace{R((2,1),N,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,3)|(2,1),N)}_{0} \cdot \left(\underbrace{R((2,1),N,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& 0 \times (5 + 0.9 \times 0) = 0 \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned}
\tag{9}
$$

We update $Q((2,1),S)$ as:

$$
\begin{aligned}
Q\left((2,1),S\right) \leftarrow & \underbrace{P((1,1)|(2,1),S)}_{0.8} \cdot \left(\underbrace{R((2,1),S,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& {\color{blue} 0.8 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(2,1),S)}_{0} \cdot \left(\underbrace{R((2,1),S,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(2,1),S)}_{0} \cdot \left(\underbrace{R((2,1),S,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(2,1),S)}_{0.1} \cdot \left(\underbrace{R((2,1),S,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(2,1),S)}_{0.1} \cdot \left(\underbrace{R((2,1),S,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(2,1),S)}_{0} \cdot \left(\underbrace{R((2,1),S,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned} \tag{10}
$$

We update $Q((2,1),W)$ as:

$$
\begin{aligned}
Q\left((2,1),W\right) \leftarrow & \underbrace{P((1,1)|(2,1),W)}_{0.1} \cdot \left(\underbrace{R((2,1),W,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(2,1),W)}_{0} \cdot \left(\underbrace{R((2,1),W,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(2,1),W)}_{0} \cdot \left(\underbrace{R((2,1),W,(1,3))}_{-5} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(2,1),W)}_{0.9} \cdot \left(\underbrace{R((2,1),W,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& {\color{blue} 0.9 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(2,1),W)}_{0} \cdot \left(\underbrace{R((2,1),W,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(2,1),W)}_{0} \cdot \left(\underbrace{R((2,1),W,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned} \tag{11}
$$

We update $Q((2,1),E)$ as:

$$
\begin{aligned}
Q\left((2,1),E\right) \leftarrow & \underbrace{P((1,1)|(2,1),E)}_{0.1} \cdot \left(\underbrace{R((2,1),E,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,2)|(2,1),E)}_{0} \cdot \left(\underbrace{R((2,1),E,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,3)|(2,1),E)}_{0} \cdot \left(\underbrace{R((2,1),E,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& 0 \times (-5 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,1)|(2,1),E)}_{0.1} \cdot \left(\underbrace{R((2,1),E,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,2)|(2,1),E)}_{0.8} \cdot \left(\underbrace{R((2,1),E,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& 0.8 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,3)|(2,1),E)}_{0} \cdot \left(\underbrace{R((2,1),E,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& 0 \times (5 + 0.9 \times 0) = 0 \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned}
\tag{12}
$$

We update $Q((2,2),N)$ as:

$$
\begin{aligned}
Q\left((2,2),N\right) \leftarrow & \underbrace{P((1,1)|(2,2),N)}_{0} \cdot \left(\underbrace{R((2,2),N,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,2)|(2,2),N)}_{0} \cdot \left(\underbrace{R((2,2),N,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,3)|(2,2),N)}_{0} \cdot \left(\underbrace{R((2,2),N,(1,3))}_{\text{-5}} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& 0 \times (-5 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,1)|(2,2),N)}_{0.1} \cdot \left(\underbrace{R((2,2),N,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,2)|(2,2),N)}_{0.8} \cdot \left(\underbrace{R((2,2),N,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& 0.8 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,3)|(2,2),N)}_{0.1} \cdot \left(\underbrace{R((2,2),N,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& 0.1 \times (5 + 0.9 \times 0) = 0.5 \\
& = 0 + 0 + 0 + 0 + 0 + 0.5 = 0.5
\end{aligned}
\tag{13}
$$

We update $Q((2,2),S)$ as:

$$\begin{aligned}
Q\left((2,2),S\right) \leftarrow & \underbrace{P((1,1)|(2,2),S)}_{0} \cdot \left(\underbrace{R((2,2),S,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,2)|(2,2),S)}_{0.8} \cdot \left(\underbrace{R((2,2),S,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& 0.8 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((1,3)|(2,2),S)}_{0} \cdot \left(\underbrace{R((2,2),S,(1,3))}_{-5} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& 0 \times (-5 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,1)|(2,2),S)}_{0.1} \cdot \left(\underbrace{R((2,2),S,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& 0.1 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,2)|(2,2),S)}_{0} \cdot \left(\underbrace{R((2,2),S,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& 0 \times (0 + 0.9 \times 0) = 0 \\
& + \underbrace{P((2,3)|(2,2),S)}_{0.1} \cdot \left(\underbrace{R((2,2),S,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& 0.1 \times (5 + 0.9 \times 0) = 0.5 \\
& = 0 + 0 + 0 + 0 + 0 + 0.5 = 0.5
\end{aligned} \tag{14}$$

We update $Q((2,2),W)$ as:

$$
\begin{aligned}
Q\left((2,2),W\right) \leftarrow & \underbrace{P((1,1)|(2,2),W)}_{0} \cdot \left(\underbrace{R((2,2),W,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0}\right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(2,2),W)}_{0.1} \cdot \left(\underbrace{R((2,2),W,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0}\right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(2,2),W)}_{0} \cdot \left(\underbrace{R((2,2),W,(1,3))}_{-5} + 0.9 \cdot \underbrace{V((1,3)}_{0}\right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(2,2),W)}_{0.8} \cdot \left(\underbrace{R((2,2),W,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0}\right) \\
& {\color{blue} 0.8 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(2,2),W)}_{0.1} \cdot \left(\underbrace{R((2,2),W,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0}\right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(2,2),W)}_{0} \cdot \left(\underbrace{R((2,2),W,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0}\right) \\
& {\color{blue} 0 \times (5 + 0.9 \times 0) = 0} \\
& = 0 + 0 + 0 + 0 + 0 + 0 = 0
\end{aligned} \tag{15}
$$

Tutorial 3.2 (a)

We update $Q((2,2),E)$ as:

$$
\begin{aligned}
Q\left((2,2),E\right) \leftarrow & \underbrace{P((1,1)|(2,2),E)}_{0} \cdot \left(\underbrace{R((2,2),E,(1,1))}_{0} + 0.9 \cdot \underbrace{V((1,1))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,2)|(2,2),E)}_{0.1} \cdot \left(\underbrace{R((2,2),E,(1,2))}_{0} + 0.9 \cdot \underbrace{V((1,2))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((1,3)|(2,2),E)}_{0} \cdot \left(\underbrace{R((2,2),E,(1,3))}_{-5} + 0.9 \cdot \underbrace{V((1,3)}_{0} \right) \\
& {\color{blue} 0 \times (-5 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,1)|(2,2),E)}_{0} \cdot \left(\underbrace{R((2,2),E,(2,1))}_{0} + 0.9 \cdot \underbrace{V((2,1))}_{0} \right) \\
& {\color{blue} 0 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,2)|(2,2),E)}_{0.1} \cdot \left(\underbrace{R((2,2),E,(2,2))}_{0} + 0.9 \cdot \underbrace{V((2,2))}_{0} \right) \\
& {\color{blue} 0.1 \times (0 + 0.9 \times 0) = 0} \\
& + \underbrace{P((2,3)|(2,2),E)}_{0.8} \cdot \left(\underbrace{R((2,2),E,(2,3))}_{5} + 0.9 \cdot \underbrace{V((2,3))}_{0} \right) \\
& {\color{blue} 0.8 \times (5 + 0.9 \times 0) = 4} \\
& = 0 + 0 + 0 + 0 + 0 + 4 = 4
\end{aligned}
\tag{16}
$$

By applying $V(S) \leftarrow \max_a Q(S, a)$, we have:

$$
\begin{aligned}
V((1,1)) \leftarrow & \max \left\{ \underbrace{Q((1,1),N)}_{0}, \underbrace{Q((1,1),S)}_{0}, \underbrace{Q((1,1),W)}_{0}, \underbrace{Q((1,1),E)}_{0} \right\} \\
& = 0 \\
V((1,2)) \leftarrow & \max \left\{ \underbrace{Q((1,2),N)}_{\text{-0.5}}, \underbrace{Q((1,2),S)}_{\text{-0.5}}, \underbrace{Q((1,2),W)}_{0}, \underbrace{Q((1,2),E)}_{\text{-4}} \right\} \\
& = 0 \\
V((1,3)) \leftarrow & 0 \quad \text{(terminal value assumption)} \\
V((2,1)) \leftarrow & \max \left\{ \underbrace{Q((2,1),N)}_{0}, \underbrace{Q((2,1),S)}_{0}, \underbrace{Q((2,1),W)}_{0}, \underbrace{Q((2,1),E)}_{0} \right\} \\
& = 0 \\
V((2,2)) \leftarrow & \max \left\{ \underbrace{Q((2,2),N)}_{0.5}, \underbrace{Q((2,2),S)}_{0.5}, \underbrace{Q((2,2),W)}_{0}, \underbrace{Q((2,2),E)}_{4} \right\} \\
& = 4 \\
V((2,3)) \leftarrow & 0 \quad \text{(terminal value assumption)}
\end{aligned}
\tag{17}
$$

We repeat the above updates except for the updated $V((2,2)) = 4$, then we have the follows for the **second iteration**:

$$
\begin{aligned}
V((1,1)) \leftarrow & \max \left\{ \underbrace{Q((1,1),N)}_{0}, \underbrace{Q((1,1),S)}_{0}, \underbrace{Q((1,1),W)}_{0}, \underbrace{Q((1,1),E)}_{0} \right\} \\
& = 0 \\
V((1,2)) \leftarrow & \max \left\{ \underbrace{Q((1,2),N)}_{2.38}, \underbrace{Q((1,2),S)}_{-0.5}, \underbrace{Q((1,2),W)}_{0.36}, \underbrace{Q((1,2),E)}_{-3.64} \right\} \\
& = 2.38 \\
V((1,3)) \leftarrow & 0 \quad \text{(terminal value assumption)} \\
V((2,1)) \leftarrow & \max \left\{ \underbrace{Q((2,1),N)}_{0.36}, \underbrace{Q((2,1),S)}_{0.36}, \underbrace{Q((2,1),W)}_{0}, \underbrace{Q((2,1),E)}_{2.88} \right\} \\
& = 2.88 \\
V((2,2)) \leftarrow & \max \left\{ \underbrace{Q((2,2),N)}_{3.38}, \underbrace{Q((2,2),S)}_{0.5}, \underbrace{Q((2,2),W)}_{0.36}, \underbrace{Q((2,2),E)}_{4.36} \right\} \\
& = 4.36 \\
V((2,3)) \leftarrow & 0 \quad \text{(terminal value assumption)}
\end{aligned}
\tag{18}
$$