Tutorial 10

1) Utilities for flats a, b and c are:

For flat a, we have that:

$P(\mathrm{AW} \mid \mathrm{F} = \mathrm{a}) = < 0.4, 0.6 >$ and $P(\mathrm{BS}|\mathrm{F} = \mathrm{a}) = < 0.1, 0.9 >$

$P(VI \mid AW, F = a) \;= < P(VI \;=\; True \mid AW, F = a), P(VI = False \mid AW, F = a >$ so,

we can start of by computing for when VI = true:
$P(VI = True|AW, F = a) = P(vi|aw, F = a) * P(aw|F = a) + P(vi|\neg aw, F = a) * P(\neg aw|F = a) = 0.6 * 0.4 + 0.1 * 0.6 = 0.3$

and for when VI = false:

$P(VI = False|AW, F = a) = P(\neg vi|aw, F = a) * P(aw|F = a) + P(\neg vi|\neg aw, F = a) * P(\neg aw|F = a) = 0.4 * 0.4 + 0.9 * 0.6 = 0.7$

Now, we can compute the utility which would be:

U(bs, vi) = $\sum_{\mathrm{BS,V}} U(BS, VI) * P(BS) * P(VI) = U(bs, vi) * P(bs) * P(vi) + U(\neg bs, vi) * P(\neg bs) * P(vi) + U(bs, \neg vi) * P(bs) * P(\neg vi) + U(\neg bs, \neg vi) * P(\neg bs) * P(\neg vi) = 0.9 * 0.1 * 0.3 + 0.5 * 0.9 * 0.3 + 0.3 * 0.1 * 0.7 + 0.1 * 0.9 * 0.7 = 0.246.$

For flat b, we have that:

$P(\mathrm{AW} \mid \mathrm{F} = \mathrm{b}) = < 0.2, 0.8 >$ and $P(\mathrm{BS}|\mathrm{F} = \mathrm{b}) = < 0.3, 0.7 >$

$P(VI \mid AW, F = b) \;= < P(VI \;=\; True \mid AW, F = b), P(VI = False \mid AW, F = b >$ so,

we can start of by computing for when VI = true:
$P(VI = True|AW, F = b) = P(vi|aw, F = b) * P(aw|F = b) + P(vi|\neg aw, F = b) * P(\neg aw|F = b) = 0.2 * 0.2 + 0.1 * 0.8 = 0.12$

and for when VI = false:

$P(VI = False|AW, F = b) = P(\neg vi|aw, F = b) * P(aw|F = b) + P(\neg vi|\neg aw, F = b) * P(\neg aw|F = b) = 0.8 * 0.2 + 0.9 * 0.8 = 0.88$

Now, we can compute the utility which would be:

U(bs, vi) = $\sum_{\mathrm{BS,V}} U(BS, VI) * P(BS) * P(VI) = U(bs, vi) * P(bs) * P(vi) + U(\neg bs, vi) * P(\neg bs) * P(vi) + U(bs, \neg vi) * P(bs) * P(\neg vi) + U(\neg bs, \neg vi) * P(\neg bs) * P(\neg vi) = 0.9 * 0.3 * 0.12 + 0.5 * 0.7 * 0.12 + 0.3 * 0.3 * 0.88 + 0.1 * 0.7 * 0.88 = 0.215.$

For flat c, we have that:

$P(\mathrm{AW} \mid \mathrm{F} = \mathrm{c}) = < 0.4, 0.6 > and\ P(\mathrm{BS} | \mathrm{F} = \mathrm{a}) = < 0.6, 0.4 >$

$P(VI \mid AW, F = c) \ = < P(VI \ = \ True \mid AW, F = c), P(VI = False \mid AW, F = c >$ so,

we can start of by computing for when VI = true:
$P(VI = True | AW, F = c) = P(vi | aw, F = c) * P(aw | F = c) + P(vi | \neg aw, F = c) * P(\neg aw | F = c) = 0.7 * 0.4 + 0.2 * 0.6 = 0.4$

and for when VI = false:

$P(VI = False | AW, F = c) = P(\neg vi | aw, F = c) * P(aw | F = c) + P(\neg vi | \neg aw, F = c) * P(\neg aw | F = c) = 0.3 * 0.4 + 0.8 * 0.6 = 0.6$

Now, we can compute the utility which would be:

U(bs, vi) = $\sum_{\mathrm{BS,V}} U(BS, VI) * P(BS) * P(VI) = U(bs, vi) * P(bs) * P(vi) + U(\neg bs, vi) * P(\neg bs) * P(vi) + U(bs, \neg vi) * P(bs) * P(\neg vi) + U(\neg bs, \neg vi) * P(\neg bs) * P(\neg vi) = 0.9 * 0.6 * 0.4 + 0.5 * 0.4 * 0.4 + 0.3 * 0.6 * 0.6 + 0.1 * 0.4 * 0.6 = 0.428.$

So, flat c has the highest utility.

2) N tends to grow quickly as γ becomes closer to 1. N tends to vary with γ for the different values of the ratio $\epsilon / R_{\mathrm{max}}$ due to the fast convergence and N doesn't depend on the ratio $\epsilon / R_{\mathrm{max}}$ that much. If γ is small then a fast convergence is obtained, however this could mean the agent misses long term effects of its actions.

3) N = $\frac{log\left(\frac{2*0.45}{0.01(1-0.1)}\right)}{log(1/0.1)} = \frac{log\left(\frac{0.9}{0.009}\right)}{log(10)}$ = $\frac{log(100)}{1} = 2.$

4) So, for the second step the utility vector is $U_1 = < -0.1, -1, -0.1, 1 >$, by placing these numbers inside the Bellman update we get:

$$\mathrm{U}_2'(s_1) = R(s_1) + \gamma max_a \begin{Bmatrix} 0.7U_1(s_3) + 0.3U_1(s_2) & \text{(up)} \\ 0.7U_1(s_2) + 0.3U_1(s_1) & \text{(right)} \\ 1U_1(s_1) & \text{(down)} \\ 0.7U_1(s_1) + 0.3U_1(s_3) & \text{(left)} \end{Bmatrix} = \text{-}0.1 + 0.1 max_a \begin{Bmatrix} 0.7(\text{-}0.1) + 0.3(\text{-}1) & \text{(up)} \\ 0.7(\text{-}1) + 0.3(\text{-}0.1) & \text{(right)} \\ 1(\text{-}0.1) & \text{(down)} \\ 0.7(\text{-}0.1) + 0.3(\text{-}0.1) & \text{(left)} \end{Bmatrix} = \text{-}0.1 + 0.1 max_a \begin{Bmatrix} \text{-}0.37 & \text{(up)} \\ \text{-}0.73 & \text{(right)} \\ \text{-}0.1 & \text{(down)} \\ \text{-}0.1 & \text{(left)} \end{Bmatrix} = \text{-}0.11.$$