H. Ayhan

Solutions to Homework 6

1. Define the state space $\mathcal{S} = \{1, 2, 3\}$, where $1 = good$, $2 = fair$, and $3 = poor$. Define the action space $\mathcal{A} = \{0, 1\}$, where $0 = Do\ Nothing$, $1 = Fertilize$. Recall that the expected immediate rewards we calculated in class are given by

$$
\begin{aligned}
r(1,0) &= 7 \times 0.2 + 6 \times 0.5 + 3 \times 0.3 = 5.3 \\
r(2,0) &= 5 \times 0.5 + 1 \times 0.5 = 3 \\
r(3,0) &= -1 \times 1 = -1 \\
r(1,1) &= 6 \times 0.3 + 5 \times 0.6 - 1 \times 0.1 = 4.7 \\
r(2,1) &= 7 \times 0.1 + 4 \times 0.6 = 3.1 \\
r(3,1) &= 6 \times 0.05 + 3 \times 0.4 - 2 \times 0.55 = 0.4
\end{aligned}
$$

And the transition matrices under each action are respectively given by:

$$
P_0 = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad P_1 = \begin{pmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{pmatrix}
$$

Then taking $\alpha = 0.8$ and $b(1) = b(2) = b(3) = 1/3$ we get that the LP primal is given by:

$$
\text{Minimize } \frac{1}{3}(v(1) + v(2) + v(3))
$$

subject to

$$
\begin{aligned}
v(1) - 0.8(0.2v(1) + 0.5v(2) + 0.3v(3)) &\geq 5.3 \\
v(2) - 0.8(0.5v(2) + 0.5v(3)) &\geq 3 \\
v(3) - 0.8(1v(3)) &\geq -1 \\
v(1) - 0.8(0.3v(1) + 0.6v(2) + 0.1v(3)) &\geq 4.7 \\
v(2) - 0.8(0.1v(1) + 0.6v(2) + 0.3v(3)) &\geq 3.1 \\
v(3) - 0.8(0.05v(1) + 0.4v(2) + 0.55v(3)) &\geq 0.4
\end{aligned}
$$

And the LP dual is

$$
\text{Maximize } 5.3 * x(1,0) + 3.0 * x(2,0) + (-1 * x(3,0)) + 4.7 * x(1,1) + 3.1 * x(2,1) + .4 * x(3,1)
$$

subject to

$$
\begin{aligned}
x(1,1) + x(1,0) - 0.8 * (.3 * x(1,1) + .1 * x(2,1) + .2 * x(1,0) + .05 * x(3,1)) &= 1/3 \\
x(2,1) + x(2,0) - 0.8 * (.5 * x(1,0) + .5 * x(2,0) + .6 * x(1,1) + .6 * x(2,1) + .4 * x(3,1)) &= 1/3 \\
x(3,1) + x(3,0) - 0.8 * (.3 * x(1,0) + .5 * x(2,0) + 1 * x(3,0) + .1 * x(1,1) + .3 * x(2,1) + .55 * x(3,1)) &= 1/3 \\
x(s,a) &\geq 0 \quad \forall s \in S, a \in A
\end{aligned}
$$

We code the LP dual into Xpress. The code follows

```
uses "mmxprs";

declarations
        states=1..3
        actions=0..1
        x :array(states, actions) of mpvar
end-declarations

obj := 5.3*x(1,0)+3.0*x(2,0)+(-1*x(3,0))+4.7*x(1,1)+3.1*x(2,1)+.4*x(3,1)

c1 := x(1,1)+x(1,0)-0.8*(.3*x(1,1)+.1*x(2,1)+.2*x(1,0)+.05*x(3,1))=1/3
c2 := x(2,1)+x(2,0)-0.8*(.5*x(1,0)+.5*x(2,0)+.6*x(1,1)+.6*x(2,1)+.4*x(3,1))=1/3
c3 := x(3,1)+x(3,0)-0.8*(.3*x(1,0)+.5*x(2,0)+1*x(3,0)+.1*x(1,1)+.3*x(2,1)+.55*x(3,1))=1/3


writeln("Begin running model")
maximize (obj)
writeln("Optimal value =", getsol(obj))
forall (s in states,a in actions)
        writeln("x(",s,",",a,")=",getsol(x(s,a)))
writeln("End running model")

end-model
```

The results are:

```
Begin running model
Optimal value =12.014
x(1,0)=0
x(1,1)=0.789263
x(2,0)=0
x(2,1)=2.45192
x(3,0)=0
x(3,1)=1.75881
End running model
```

From this we conclude the optimal policy is $d^*(1) = 1, d^*(2) = 1, d^*(3) = 1$. That is, always fertilize.

2. For simplicity we will use the shorthand $s_1 = 1$, $s_2 = 2$

   (a) We set $\alpha = 0.5$, $n = 0$ and do the value iteration. Let us arbitrarily take $V_0(s_1) = V_0(s_2) = 0$. And for good measure let's take $\epsilon = 10^{-5}$.

$$V_1(1) = \max\{1\ ,\ 0\} = 1$$
$$V_1(2) = \max\{2\} = 2$$

   We calculate the stopping condition and get

$$\max\{|1-0|, |2-0|\} = 2 \not< 3.33 \times 10^-6 = 10^{-5}\frac{1-0.6}{2\times 0.6}$$

   so we continue, set $n = 1$ and carry on. I used the same code from the previous homework and got:

```
    Solver set to Value Iter. Solver  (Disc)
 2 states found.

    Max difference from previous value = 2.0
    Max difference from previous value = 1.0
    Max difference from previous value = 0.5
    Max difference from previous value = 0.25
    Max difference from previous value = 0.125
    Max difference from previous value = 0.0625
```

```
      Max difference from previous value = 0.03125
      Max difference from previous value = 0.015625
      Max difference from previous value = 0.0078125
      Max difference from previous value = 0.00390625
      Max difference from previous value = 0.001953125
      Max difference from previous value = 9.765625E-4
      Max difference from previous value = 4.8828125E-4
      Max difference from previous value = 2.44140625E-4
      Max difference from previous value = 1.220703125E-4
      Max difference from previous value = 6.103515625E-5
      Max difference from previous value = 3.0517578125E-5
Value Iter. Solver  (Disc)
********* Best Policy *********

In every stage do:
STATE       ------> ACTION
(1)         ------> (1)
(2)         ------> (1)
Value Function:
(1)           :      -2.00
(2)           :      -4.00
17 iterations
```

In our notation that is $d^*(1) = a_{1,1}, d^*(2) = a_{2,1}$.

(b) We now set $\alpha = 0.7$, $n = 0$ and do the value iteration again. Again take $V_0(s_1) = V_0(s_2) = 0$ and $\epsilon = 10^{-5}$.

$$V_1(1) = \max\{1\ ,\ 0\} = 1$$
$$V_1(2) = \max\{2\} = 2$$

We calculate the stopping condition and get

$$\max\{|1-0|, |2-0|\} = 2 \not< 3.33 \times 10^-6 = 10^{-5}\frac{1-0.6}{2\times 0.6}$$

so we continue, set $n = 1$ and carry on. I used the same code as before and got:

```
      Solver set to Value Iter. Solver  (Disc)
  2 states found.

      Max difference from previous value = 2.0
      Max difference from previous value = 1.4
      Max difference from previous value = 0.98
      Max difference from previous value = 0.6859999999999999
      Max difference from previous value = 0.48019999999999996
...
      Max difference from previous value = 3.1555076406952765E-5
      Max difference from previous value = 2.2088553484955753E-5
      Max difference from previous value = 1.546198743973548E-5
Value Iter. Solver  (Disc)
********* Best Policy *********

In every stage do:
STATE       ------> ACTION
(1)         ------> (2)
(2)         ------> (1)
Value Function:
(1)           :      -4.67
(2)           :      -6.67
34 iterations
```

In our notation that is $d^*(1) = a_{1,2}, d^*(2) = a_{2,1}$. Note I cut out a lot of the intermediate iterations in the interest of space.

(c) Let us write the primal and dual LP's in terms of a general α. First the primal. We take $b(1) = b(2) = 1/2$. Then the LP primal is given by:

$$\text{Minimize} \;\; \frac{1}{2}\left(v(1) + v(2)\right)$$

subject to

$$\begin{aligned} v(1) - \alpha(1v(1)) &\geq 1 \\ v(1) - \alpha(1v(2)) &\geq 0 \\ v(2) - \alpha(1v(2)) &\geq 2 \end{aligned}$$

And the LP dual is

$$\text{Maximize} \;\; 1 * x(1, a_{1,1}) + 2 * x(2, a_{2,1})$$

subject to

$$\begin{aligned} x(1, a_{1,1}) + x(1, a_{1,2}) - \alpha * (x(1, a_{1,1})) &= 1/2 \\ x(2, a_{2,1}) - \alpha * (x(1, a_{1,2}) + x(2, a_{2,1})) &= 1/2 \\ x(s, a) \geq 0 \;\; \forall s \in S, a \in A \end{aligned}$$

We code this into Xpress using $\alpha = 0.5$, $a_{1,1} = 1$, $a_{1,2} = 2$ and $a_{2,1} = 1$. The code follows

```
model Ex2
uses "mmxprs";

declarations
        states=1..2
        actions=1..2
        x :array(states, actions) of mpvar
end-declarations

obj := 1*x(1,1)+2*x(2,1)

const1 := x(1,1)+x(1,2)-0.5*(x(1,1))=(1/2)
const2 := x(2,1)-0.5*(x(1,2)+x(2,1))=(1/2)

const3:=        x(2,2)=0

writeln("Begin running model")
maximize (obj)
writeln("Optimal value =", getsol(obj))
forall (s in states,a in actions)
        writeln("x(",s,",",a,")=",getsol(x(s,a)))
writeln("End running model")

end-model
```

We run the code and get

```
Begin running model
Optimal value =3
x(1,1)=1
x(1,2)=0
x(2,1)=1
x(2,2)=0
End running model
```

So we conclude when $\alpha = 0.5$ the optimal policy is $d^*(1) = a_{1,1}, d^*(2) = a_{2,1}$.

We change α to 0.7 in the code above and run it again and get.

```
Begin running model
Optimal value =5.66667
x(1,1)=0
x(1,2)=0.5
x(2,1)=2.83333
x(2,2)=0
End running model
```

So we conclude when $\alpha = 0.7$ the optimal policy is $d^*(1) = a_{1,2}, d^*(2) = a_{2,1}$.

Note that as expected, the optimal solution matches for both methods.