Homework #18
Due by 7AM, Monday, April 20

Using the maintenance problem given in the example, we wish to maximize profits. Profits are determined by revenue minus costs, where revenue is a function of the state but not of the decision and is given by the function $\mathbf{r} = (900, 400, 450, 750)^T$. The costs and transition probabilities are as presented in the example.

1. Give the profit functions $\mathbf{g}_k$ for $k \in E$.

   $\mathbf{g}_1 = (800, 275, 300, 250)^T$ and $\mathbf{g}_2 = (600, 75, 100, 150)^T$

2. Using a discount factor $\alpha=0.95$, we would like to verify that

   $$\mathbf{v}^\alpha = (8651.88, 8199.73, 8233.37, 8402.65)^T$$

   To verify that it is optimal, you would use the fixed point property and check and calculate $v^\alpha(a)$, $v^\alpha(b)$, $v^\alpha(c)$, $v^\alpha(d)$. To save time, only check $v^\alpha(d)$. As you calculate $v^\alpha(d)$, you assume $\mathbf{v}^\alpha$ is optimal and use it on the right-hand side of the equation in Property (1) and finally obtain that $v^\alpha(d)$ is the maximum of two numbers. What is *smaller* of those two numbers?

   $v^\alpha(a) = \max\{ 250 + 0.95(0.8, 0.1, 0, 0.1)\,\mathbf{v}^\alpha \;;\; 150 + 0.95(0.9, 0.1, 0, 0)\,\mathbf{v}^\alpha \}$

   $= \max\{8402.65 \;;\; 8326.33\} = 8402.65$

3. Using the policy improvement algorithm, find the policy that maximizes the total discounted profit if the discount factor is such that \$1 today is worth \$1.12 after one time period.

   a. What is the discount factor? Round to two decimals to the right of the decimal, then use that same value (i.e., the rounded value) as you do the remainder of the problem. $\alpha = 1/1.12$ and rounded yields 0.89

   b. What is the optimum policy? $\mathbf{a}_0 = (1,1,1,1) \rightarrow \mathbf{v} = (3993, 3482, 3573, 3736)$

      $a_1(a) = \text{maxarg}\{3993, 3980\}$, $a_1(b) = \text{maxarg}\{3482, 3535\}$
      $a_1(c) = \text{maxarg}\{3573, 3563\}$, $a_1(d) = \text{maxarg}\{3736, 3659\}$
      $\mathbf{a}_1 = (1,2,1,1) \rightarrow \mathbf{v} = (4068, 3605, 3640, 3806)$.

      $a_2(a) = \text{maxarg}\{4068, 4059\}$, $a_2(b) = \text{maxarg}\{3553, 3605\}$
      $a_2(c) = \text{maxarg}\{3640, 3638\}$, $a_2(d) = \text{maxarg}\{3806, 3730\}$
      Thus, $\mathbf{a}_2 = (1,2,1,1)$ and since $\mathbf{a}_2 = \mathbf{a}_1$ we are finished.

   c. What is the expected total discounted profit when the optimum policy is used assuming that you start with a new machine (i.e., start in state a). 4068

   Note: using value iteration, it takes around 30 steps to have a good answer, with policy iteration, you should have the final answer after n=2.