Homework #1

Question 2

1. Formulation of the problem as a finite horizon decision problem:

   State Space: $S = \{(i,j) \ \mid \ i \in \{1,\dots,M\},\ j \in \{1,\dots,N\}\}$, overall $M \cdot N$ states.

   Action Space: $A = \{Up, Right\}$

   Definitions of helper function f_t which return the immediate reward for a given state and action and the cumulative cost function $C\colon S \to \mathbb{N}$:

   $$f_t(s_t, a_t) = \begin{cases} 1 & \text{there is a cheese in } s_t \\ 0 & \text{there isn't cheese in } s_t \end{cases}$$

   $$f_T(s_T) = \begin{cases} 1 & \text{there is a cheese in } (M,N) \text{ position} \\ 0 & \text{there isn't cheese in } (M,N) \text{ position} \end{cases}$$

   $$C(s_0, a_0, \dots, s_{T-1}, a_{T-1}, s_T) = \sum_{i=1}^{T-1} f_t(s_t, a_t) + f_T(s_T)$$

2. When $N = 8, M = 5$ the game is always over after 11 steps. In order that the mouse will be at the top right corner the mouse needs to take 4 steps UP and 7 steps RIGHT, or under general settings, M-1 steps UP and N-1 steps RIGHT.

   Overall, the horizon will be $M - 1 + N = M + N - 2$.

3. When $M = 2$ the number of trajectories is $\binom{N}{1} = N$.

   This is intuitive because when $M = 2$ we can go UP just one time. Given we choose the place the mouse will go UP the rest of the steps are RIGHT.

   When $M = N$, we need to go UP N steps, out of $M + N - 2$ possible steps.

   Once we choose when we will go UP the rest of the steps have to be RIGHT.

   Overall, the number of trajectories is $\binom{N+N-2}{N-1} = O(N^N)$

   Under general settings the number of trajectories is $\binom{N+M-2}{M-1} = O(N^M)$.

4.

(a) if both mouses ignore each other's existence and act "optimal" with respect to the original problem one of the mice will eat a lot of cheese and the other one will starve.

The action space is similar to section 1, since the actions the mouses can take didn't change, but because each mouse moves independently of the other mouse, we will get 4 different actions: $\mathrm{UP}_1, \mathrm{RIGHT}_1, \mathrm{UP}_2, \mathrm{RIGHT}_2$.

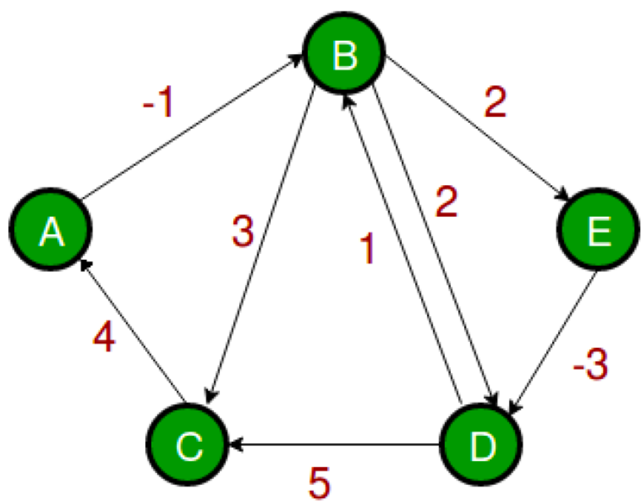
The state space is changed since we need to address two mice moving and not only one, hence: $S = \left\{(i_1, j_1, i_2, j_2) \;\middle|\; i_1, i_2 \in \{1, \dots, M\},\ j_1, j_2 \in \{1, \dots, N\}\right\}$

The number of states is N^2M^2 .

(b) There are $N^K M^K$ states and 2^K actions, 2 option for each mouse (up or right).

Question 4

The following graph describes a Deterministic Decision Processes with an initial state $s = A$.



1. Using dynamic programming

	$d_0[\cdot]$	$d_1[\cdot]$	$d_2[\cdot]$	$d_3[\cdot]$	$d_4[\cdot]$	$d_5[\cdot]$
$A \to A$	0	∞	∞	6	10	7
$A \to B$	∞	-1	∞	2	-1	5
$A \to C$	∞	∞	2	6	3	2
$A \to D$	∞	∞	1	-2	4	1
$A \to E$	∞	∞	1	∞	4	1

2. Karp's Algorithm: $\mu^* = \min_{v \in V} \max_{0 \le k \le n-1} \frac{d_n(v) - d_k(v)}{n-k}$

$$\min_{v \in V} \left\{ \frac{7-0}{5-0}, \frac{5-(-1)}{5-4}, \frac{2-2}{5-2}, \frac{1-(-2)}{5-3}, \frac{1-1}{5-2} \right\} = \min_{v \in V} \left\{ \frac{7}{5}, \frac{6}{1}, 0, \frac{3}{2}, 0 \right\} = 0$$

3. The optimal average cost of this DDP is 0
   the cycle is $B \to E \to D \to B$

Programming Part

Question 2