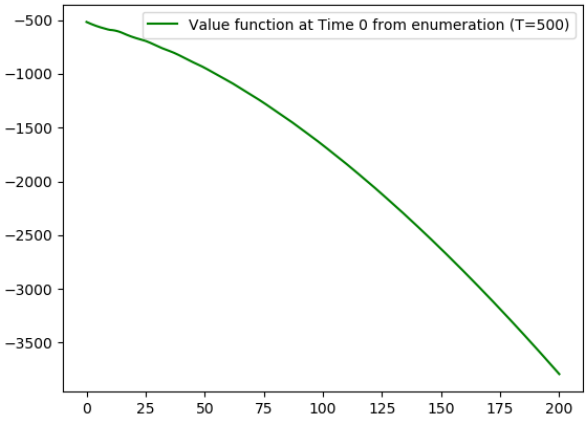
Reinforcement Learning HW-1

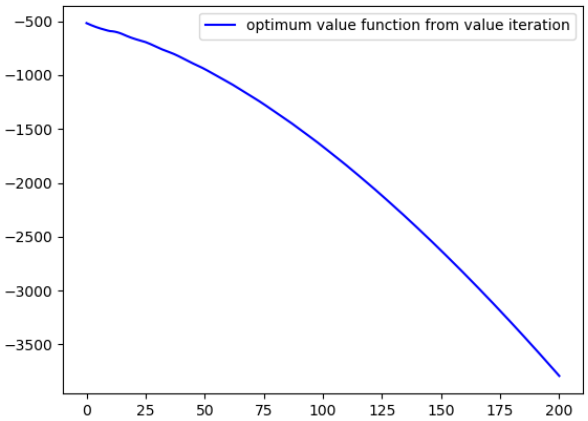
Name: Li, Jieda
NetID: jlg7773

Problem 1: Please see plots for problem 1(a), 1(b) and 1(c) below.
[Reference Code]

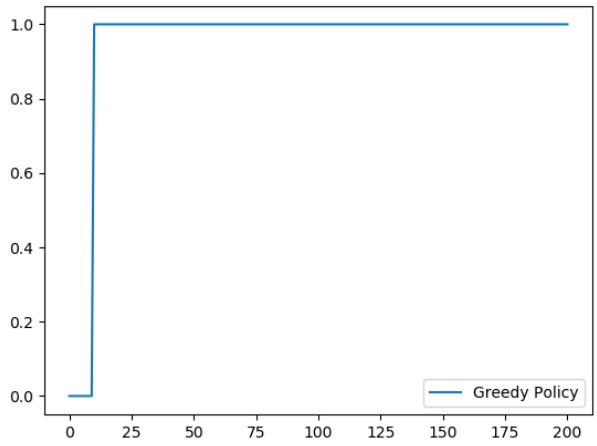
1(a) Value function at Time 0 (from enumeration method with T=500)



1(b) Value function from value iteration



1(c) Optimum policy from policy iteration



Problem 2:

[Reference Code]

Modeling details / assumptions:

- State is now defined as a matrix, shape 101 *5
- Value function is defined as a vector, length is 101**5 = 10,510,100,501
- Total number of states are 101**5 = 10,510,100,501
- Each unique state is represented by a vector of 5; Such as, S = [10 5 6 67 25]
- At one time point, arriving customer is represented by a vector of 5, such as [1 5 3 4 2]
- **When a bus is dispatched, it will always take 6 people from each class; If a class has less than 6 people, it will just take whatever number of customers in that class.**
  For example, current state S = [10 5 6 67 25], when a shuttle is dispatched, next state will be [10 5 6 67 25] – [6 5 6 6 6] → [4 0 0 61 19]
- Each column of the value function matrix corresponds to a specific customer type
- Maximum number of customers per class type waiting is 100
- Capacity of shuttle is 30
- Cost of each remaining customer is ch = {1 1.5 2 2.5 3}

Mathematical approximations used to reduce computation load:

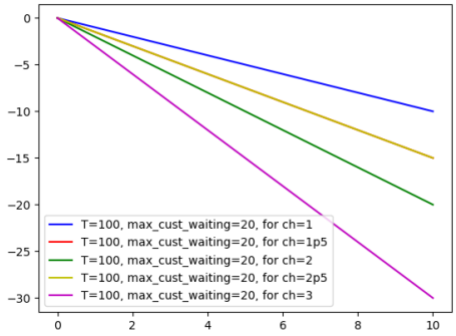
- Calculation of expected total future rewards per action (expectation over all next states) is replaced by sampling of next state (one sample).

Experiment Results:

- **Attempt 1:**
  Run with max number of customer waiting = 100, total states = 10,510,100,501
  Result: **Ipython killed the process reporting "kill 9".** (consuming too much memory)

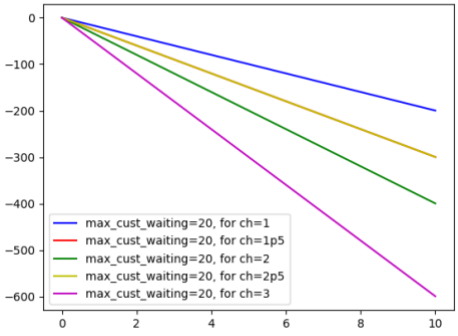
- **Attempt 2:**
  Run with ***reduced state space***:
  reduce max customer waiting from 100 to 10, total states = 4,084,101
  Result: show below the value function from enumeration and value iteration
  X: number of customers in each class (max=20)

2(a) Value function from enumeration T=100



2(b) Value function from value iteration

- **takes T= 134, about half an hour, to converge with theta = 0.1**



- **Comment:** with original specified state space, it is not possible to run this algorithm within the capability of my laptop. The vector for value functions are too big for memory and the loop to loop through all states are 10,510,100,501, **memory consumption and time needed to run** are both beyond the capability of my laptop.