Final project

We consider a rectangular (NY-like) city of 50×50 crossings. There are thus 2500 states and 4 actions. You can consider $\gamma = 0.99$.

The goal is to find the shortest route to (1,10) where r=1 (and which is a terminal state).
Due to congestion, transition probability per road segment vary iid $\in \{0.1, ..., 1\}$ each with probability $1/10$.
(1) Generate a realization of the congestion.
(1) Think of a simple heuristic that solves this problem and implement it
In the case of direct access to the environment parameters, compute the fastest route
(2) Via solving a system of equations
(2) Via dynamic programming

Now consider that you only have access to the model as a black-box (an agent receives info on the reward and the next state after taking an action in a given state).
(2) Use Q-learning with ϵ-greedy and initial Q-values 0. You can choose ϵ, and the initial state (or a distribution of initial states) of the agent (once in a terminal state the agent starts a new episode).
(1) Think of a way to modify the reward function in a way to ease learning with Q-learning while introducing a limited bias (still with ϵ-greedy and initial Q-values 0)
(1) Compare in a useful way the outcomes of the different approaches
Submit code (python) and explanation (2 pages pdf)