Surname, Name, ID number: ______________________________

Autonomous and Mobile Robotics M

27 January 2022 - Theory

Some questions may have more than one correct answers: for each question, indicate all the correct answers.

1. The configuration space of a bicycle mobile robot is:
   - ○ $[x\ y\ \theta]^T \in \mathbb{R}^2 \times \mathbb{S}$
   - ○ $[x\ y\ \theta\ \gamma]^T \in \mathbb{R}^4$
   - ⊗ $[x\ y\ \theta\ \gamma]^T \in \mathbb{R}^2 \times \mathbb{S}^2$
2. A *non-holonomic* constraint:
   - ⊗ cannot be fully integrated
   - ○ can be written in the configuration space
   - ⊗ does not restrict the space of configurations but the instant robot mobility
3. In a trajectory following problem, the robot must asymptotically perform a desired trajectory
   - ⊗ that depends on a free parameter s
   - ○ that depends on time t
   - ○ that must be represented by a polynomial function
4. In map-based navigation, the robot:
   - ⊗ plans the trajectory using a map of the environment
   - ○ must update the planned path on the based of sensor information
   - ○ navigates the environment on the base of the sensor information only
5. Sequential Monte-Carlo Localization:
   - ⊗ is based on the simultaneous evaluation of multiple potential robot configurations called particles;
   - ⊗ resamples the particles on the based of the weights evaluated after prediction, innovation and normalization;
   - ○ works only in case the robot configuration can be described by a Gaussian distribution.
6. The Bellman optimality equation for the action value function can be written as
   - ○ $q_*(s,a) = \max v_\pi(s)$
   - ⊗ $q_*(s,a) = \mathbb{E}_*[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a]$
   - ○ $q_*(s,a) = p(s', r|s,a)\,[r + \gamma \max_{a' \in \mathcal{A}} q_*(s', a')]$
7. Monte-Carlo reinforcement learning:
   - ○ can be applied to non-episodic tasks
   - ○ requires the knowledge of the reward model
   - ⊗ does not require the knowledge of the transition model
8. Under which hypotheses Monte-Carlo reinforcement learning methods converge to optimal value function?
   - ⊗ infinite episodes
   - ○ deterministic policy
   - ⊗ exploring starts
9. The λ-return is defined as:
   - ⊗ $G_t^\lambda = (1-\lambda) \sum_{n=1}^{\infty} \lambda^{n-1} G_t^{(n)}$
   - ○ $G_t^\lambda = \sum_{k=0}^{\infty} \lambda^k R_{t+k+1}$
   - ○ $G_t^\lambda = R_{t+1} + \lambda V(S_{t+1})$
10. In value function approximation by stochastic gradient descent, the parameter vector update is defined as:
    - ○ $\Delta \mathbf{w} = -\frac{1}{2} \alpha \nabla_{\mathbf{w}} J(\mathbf{w})$
    - ⊗ $\Delta \mathbf{w} = \alpha\,(v_\pi(S) - \hat{v}(S,\mathbf{w})) \nabla_{\mathbf{w}} \hat{v}(S,\mathbf{w})$
    - ○ $\Delta \mathbf{w} = \alpha\,\mathbb{E}_\pi\,[(v_\pi(S) - \hat{v}(S,\mathbf{w})) \nabla_{\mathbf{w}} \hat{v}(S,\mathbf{w})]$

Autonomous and Mobile Robotics M

27 January 2022 - Exercise

The student is asked to solve the following problem.

Let us consider a fully observable and deterministic environment with 5 states $s_{\{1,\dots,5\}}$.

s_1	s_2	s_3	s_4	s_5

- Action set : {TryLeft, TryRight}
- Rewards:
  - +1 in state s_1
  - -1 in state s_3
  - +2 in state s_5
  - 0 in all other states
- Transition model:
  - $p(s_1|s_1, \text{TryLeft}) = p(s_5|s_5, \text{TryRight}) = 1$
  - $p(s_1|s_1, \text{TryRight}) = p(s_2|s_1, \text{TryRight}) = 0.5$
  - $p(s_1|s_2, \text{TryLeft}) = p(s_2|s_2, \text{TryLeft}) = 0.5$
  - $p(s_2|s_2, \text{TryRight}) = p(s_3|s_2, \text{TryRight}) = 0.5$
  - ...
- Policy: $\pi(\text{TryLeft}|s_{\{1,\dots,5\}}) = \pi(\text{TryRight}|s_{\{1,\dots,5\}}) = 0.5$
- Discount factor $\gamma = 0.9$

Starting from an arbitrary initialisation of the state value function, compute the first iteration of the state value function evaluation provided by a Dynamic Programming algorithm assuming the random policy π.

$v_\pi(s_1)$	$v_\pi(s_2)$	$v_\pi(s_3)$	$v_\pi(s_4)$	$v_\pi(s_5)$

Solution:

The state value function is initialized to 0 for all the states.

$v_\pi(s_1)$	$v_\pi(s_2)$	$v_\pi(s_3)$	$v_\pi(s_4)$	$v_\pi(s_5)$
0.75	0	-0.5	0.25	1.5