XY1234 Course Name

Exam -- Jan 1970

Division of Decision and Control Systems
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology

Re-exam (omtentamen), January 1st, 1970, kl 00.00 - 05.00

Aids. Slides of the lectures (not exercises), lecture notes (summary.pdf), mathematical tables.

Observe. Do not treat more than one problem on each page. Each step in your solutions must be motivated. Write a clear answer to each question. Write name and personal number on each page. Please only use one side of each sheet. Mark the total number of pages on the cover.

Grading.
Grade A: ≥ 43 Grade B: ≥ 38
Grade C: ≥ 33 Grade D: ≥ 28
Grade E: ≥ 23 Grade Fx: ≥ 21

Responsible. John Doe aaaxxxyyzz

Results. Posted no later than January 15th, 1970

Good luck!

Problem 1 - Quiz

(a) Why can't we use the Policy Gradient approach for off-policy learning? [1 pt]

(b) Consider the following problem: in each round you choose a scalar θ and you observe a random variable $f(\theta)$ such that $\mathbb{E}[f(\theta)] = g(\theta)$. Which technique would you use to solve in θ the equation $g(\theta) - \alpha = 0$? (for some given α) [1 pt]

(c) What is the complexity (number of operations) of solving the Bellman's equations in a finite time-horizon MDP with S states, A actions, and time-horizon T ? [1 pt]

(d) Is the SARSA algorithm based on the stochastic approximation algorithm or the stochastic gradient algorithm? [1 pt]

(e) Is the Q-learning algorithm with function approximation based on the stochastic approximation algorithm or the stochastic gradient algorithm? [1 pt]

(f) In SARSA, we propose to use ε-greedy policy with a value of ε decreasing over time. More precisely, we select $\varepsilon_t = \frac{1}{t^2}$. The algorithm does not seem to converge. Can you explain why? [1 pt]

(g) In actor-critic algorithms, how many parameters do we need to update? What do they correspond to? [1 pt]

(h) Suppose we take the step-size $\alpha_t = \frac{1}{\log(t)}$ in the Q-learning algorithm. Are the iterates guaranteed to converge almost surely to the true Q-function? [1 pt]

(i) Let $X_1, X_2, ...$ be an homogenous Markov chain with finite state space. Is the reverse process starting at time N also a Markov chain? (The reverse process is $(X_N, X_{N-1}, ..., X_1)$) [2 pts]