SYLLABUS

<table>
<tr><td colspan="3">Course Type</td><td colspan="5">Elective</td></tr>
<tr><td colspan="3">Course Code</td><td colspan="5">EE619</td></tr>
<tr><td colspan="2">Course Name</td><td>Korean</td><td colspan="5">강화학습이론</td></tr>
<tr><td colspan="2">Course Name</td><td>English</td><td colspan="5">Mathematical Foundations of Reinforcement Learning</td></tr>
<tr><td colspan="3">Name</td><td colspan="5">정송</td></tr>
<tr><td colspan="3">Lecture:Exp.:Credit(Homework)</td><td colspan="5">3:0:3(6)</td></tr>
<tr><td colspan="4">Mutually Recognized Course(BS/MS) : (X)</td><td colspan="4">Term : Spring</td></tr>
<tr><td>Class website</td><td colspan="3">http://netsys.kaist.ac.kr/lecture/EE619/EE619_2019.html</td><td colspan="2">Classromm/Time</td><td colspan="2">N1 #201/Wed 19:00~21:30</td></tr>
<tr><td rowspan="5">TA</td><td>Name</td><td>Office Hours</td><td>Office</td><td colspan="2">Tel</td><td colspan="2">Email</td></tr>
<tr><td>Yongsik Lee</td><td>TBA</td><td>N1 #918</td><td colspan="2">042-350-5473</td><td colspan="2">dldydtlr93@kaist.ac.kr</td></tr>
<tr><td>Keunhyung Chung</td><td>TBA</td><td>N1 #918</td><td colspan="2">042-350-5473</td><td colspan="2">kh.chung@kaist.ac.kr</td></tr>
<tr><td>Jinyeong Lee</td><td>TBA</td><td>N1 #620</td><td colspan="2">042-350-7524</td><td colspan="2">jylee120@kaist.ac.kr</td></tr>
<tr><td>Hyunwoo Jung</td><td>TBA</td><td>N1 #620</td><td colspan="2">042-350-7524</td><td colspan="2">destinylz@kaist.ac.kr</td></tr>
<tr><td colspan="2">Textbook</td><td colspan="6">Neuro-Dynamic Programming, Dimitri P. Bertsekas and John Tsitsiklis, Athena Scientific, 1996
Dynamic Programming and Optimal Control, Vol. II: Approximate Dynamic Programming, Dimitri P. Bertsekas, Athena Scientific, 2012
Abstract Dynamic Programming, Dimitri P. Bertsekas, Athena Scientific, 2013
Lecture Notes.</td></tr>
<tr><td colspan="2">Late Submission & Copy Penalty of Assignment</td><td colspan="6">Late submission will not be accepted, and no score will be given to copied submission.</td></tr>
<tr><td colspan="2">Grading Policy</td><td colspan="6">Midterm (35%) Final (35%) Homework (20%) Attendance (10%)
(Homework may include programming assignments).</td></tr>
<tr><td colspan="2">Descriptions of Courses</td><td colspan="6">The subject of this course is sequential decision making under uncertainty in a system whose evolution is influenced by decisions. The decision made at any given time depends on the state of the system and the objective is to select a decision making rule that optimizes a certain performance criterion. Such problems can be solved, in principle, using the classical methods of dynamic programming. In practice, however, the applicability of dynamic programming to many important problems is limited by the enormous size of the underlying state/action spaces as well as uncertainties in the system. “Neuro-dynamic programming” or "Reinforcement Learning" which is the term used in the Artificial Intelligence literature, uses neural networks and other approximation architectures to overcome such bottlenecks to the applicability of dynamic programming, while using Mote Carlo estimation and/or stochastic approximation to learn models or value functions of the system. The methodology allows systems to learn about their behavior through simulation, and to improve their performance through iterative reinforcement. The focus of this course is to understand the mathematical foundations of this methodology in light of the convergence, degree of suboptimality, computational complexity and sample efficiency of different algorithms.</td></tr>
</table>

•Schedule

Period	Topics	Remarks
Week 1	Markov Decision Processes and Dynamic Programming	
Week 2	Markov Decision Processes and Dynamic Programming	
Week 3	Simulation-Based Methods	
Week 4	Simulation-Based Methods	
Week 5	Value Function Approximation	
Week 6	Value Function Approximation	
Week 7	Asynchronous Stochastic Approximation and Convergence	
Week 8	Mid-term exam	
Week 9	Problem Approximation and Reduction	
Week 10	Policy Search Methods and Bayesian Optimization	
Week 11	Policy Search Methods and Bayesian Optimization	
Week 12	Multi-armed Bandit as ADP	
Week13	Multi-armed Bandit as RL	
Week 14	Online Learning and Games	
Week 15	Online Learning and Games	
Week 16	Final exam	

* Laboratory topics should be given in Remark for courses with lab hours.

Name of Professor : Sign

Chairman : Sign