Ders Bilgi Formu

Ders Adı	Kodu	Yerel Kredi	AKTS	Ders (saat/hafta)	Uygulama (saat/hafta)	Laboratuar (saat/hafta)
Derin Pekiştirmeli Öğrenme	MTM5137	3	7.5	3	0	0

Önkoşullar	Yok

Yarıyıl	Güz, Bahar

Dersin Dili	İngilizce, Türkçe
Dersin Seviyesi	Yüksek Lisans Seviyesi
Ders Kategorisi	Uzmanlık/Alan Dersleri
Dersin Veriliş Şekli	Yüz yüze

Dersi Sunan Akademik Birim	Matematik Mühendisliği Bölümü
Dersin Koordinatörü	Mert Bal
Dersi Veren(ler)	Mert Bal, Kadriye Şimşek Alan
Asistan(lar)ı	

Dersin Amacı	
Dersin İçeriği	
Opsiyonel Program Bileşenleri	Yok

Ders Öğrenim Çıktıları

Haftalık Konular ve İlgili Ön Hazırlık Çalışmaları

Hafta	Konular	Ön Hazırlık
1	Pekiştirmeli Öğrenmeye Giriş ve Markov Karar Süreçleri	Kavramsal: RL'nin supervised/unsupervised learning'den farkını inceleyin. Agent, environment, state, action, reward, return, policy ve value kavramlarını öğrenin. Matematik: Olasılık dağılımları, koşullu olasılık ve beklenen değer kavramlarını gözden geçirin. Okuma: Sutton & Barto, Bölüm 1–3. Hazırlık sorusu: Gerçek hayattan bir problemi MDP olarak nasıl formüle ederdiniz?

2	Bellman Denklemleri ve Dinamik Programlama	Kavramsal: Value function ve Q-function arasındaki farkı inceleyin. Matematik: Recursion, expectation, discount factor kavramlarını tekrar edin; Bellman operatörünün bir contraction mapping olduğu ve buradan yakınsama garantisinin nasıl geldiğini araştırın. Okuma: Sutton & Barto, Bölüm 3–4. Hazırlık: Bellman expectation ve Bellman optimality denklemleri arasındaki farkı açıklamaya çalışın. Kısa çalışma: Küçük bir MDP için value iteration'ın 2–3 iterasyonunu yapın.
3	Modelden Bağımsız Pekiştirmeli Öğrenme: Monte Carlo, Zamansal Fark Öğrenme ve Q-Öğrenme	Kavramsal: Model-based ve model-free RL farkını inceleyin. Monte Carlo ve TD learning'i karşılaştırın. Okuma: Sutton & Barto, Bölüm 5–6. Matematik: TD error ve bootstrapping kavramlarını gözden geçirin. Hazırlık: SARSA ile Q-Öğrenme'nin (Q-Learning) neden farklı politikalar üretebildiğini düşünün (on-policy vs off-policy). Kod: Basit bir Q-Öğrenme uygulamasının pseudocode'unu inceleyin.
4	Değer Tabanlı Derin Pekiştirmeli Öğrenme ve Derin Q-Network	Kavramsal: Q-Öğrenme'nin büyük state space'lerde neden zorlandığını inceleyin. Neural network ile function approximation kavramını tekrar edin. Okuma: DQN temel makalesi (Mnih et al. 2015) ve Sutton & Barto'da function approximation bölümleri. Hazırlık: Experience replay ve target network neden gerekli olabilir? Kod: PyTorch'ta basit bir MLP ve loss/backpropagation yapısını gözden geçirin.
5	Politika Gradyanı Yöntemleri	Matematik: Gradient, chain rule ve log-derivative trick'i gözden geçirin; Policy Gradient Theorem'in nasıl türetildiğine dair sezgiyi kazanmaya çalışın. Kavramsal: Value-based ve policy-based RL farkını inceleyin. Okuma: Sutton & Barto, Policy Gradient bölümü. Hazırlık: REINFORCE algoritmasının neden doğrudan policy parametrelerini güncellediğini anlamaya çalışın. Soru: Policy gradient yöntemlerinde neden yüksek variance problemi ortaya çıkabilir?

6	Aktör-Eleştirmen Yöntemleri	Ön bilgi: Policy gradient ve REINFORCE tekrar edilmeli. Kavramsal: Actor, critic, TD error ve advantage kavramlarını inceleyin. Matematik: Advantage
7	Ileri Politika Optimizasyonu ve Proksimal Politika Optimizasyonu (PPO)	Ön bilgi: Policy gradient, actor-critic ve advantage estimation tekrar edilmeli. Matematik: Importance sampling ve probability ratio kavramlarını inceleyin. Okuma: PPO temel makalesi (Schulman et al. 2017). Hazırlık: Policy'nin çok büyük adımlarla güncellenmesinin neden problem yaratabileceğini düşünün. Soru: PPO clipping mekanizması hangi problemi çözmeye çalışıyor?
8	Midterm 1 / Practice or Review	
9	Sürekli Kontrol ve Entropi Tabanlı Pekiştirmeli Öğrenme ve Keşif	Ön bilgi: Actor-critic ve policy optimization tekrar edilmeli. Kavramsal: Discrete vs continuous action spaces farkını inceleyin. DDPG, TD3 ve SAC'ın temel fikirlerine bakın. Klasik keşif (exploration) kavramlarını inceleyin: ε-greedy, optimistic initialization, intrinsic motivation/curiosity-driven exploration. Matematik: Entropy ve entropy regularization kavramlarını gözden geçirin. Hazırlık: SAC'ın neden entropy'yi objective'e dahil ettiğini ve bunun keşifle ilişkisini anlamaya çalışın. Karşılaştırma: PPO ve SAC'ın temel farklarını tablo halinde hazırlayın.
10	Veri Verimli Pekiştirmeli Öğrenme	Kavramsal: Sample efficiency kavramını ve neden RL'in genelde çok fazla veriye ihtiyaç duyduğunu inceleyin. Experience Replay ve Prioritized Experience Replay mekanizmalarına bakın. Hindsight Experience Replay (HER)'in seyrek ödül (sparse reward) problemini nasıl çözdüğünü araştırın. Okuma: Schaul et al. 2016 (Prioritized Experience Replay), Andrychowicz et al. 2017 (HER). Hazırlık: Off-policy yöntemlerin on-policy yöntemlere göre neden genelde daha veri verimli olduğunu düşünün. Tartışma: Daha fazla veri toplamak mı, mevcut veriyi daha verimli kullanmak mı?

11	Model Tabanlı Pekiştirmeli Öğrenme ve Planlama	Ön bilgi: MDP ve model-free RL tekrar edilmeli. Kavramsal: Transition model, reward model ve environment model kavramlarını inceleyin. Dyna ve MCTS hakkında temel bilgi edinin. Hazırlık: Model-based RL'in model-free RL'e göre temel avantajı nedir? Soru: Yanlış öğrenilmiş bir environment modelinin policy üzerindeki etkisi ne olabilir?
12	Offline Pekiştirmeli Öğrenme	Kavramsal: Online vs offline RL farkını inceleyin. Dataset, behavior policy, distribution shift ve extrapolation error kavramlarını öğrenin. Okuma: Levine et al. 2020 (Offline RL: Tutorial, Review, and Perspectives on Open Problems). Hazırlık: Agent'ın environment ile yeni etkileşime giremediği bir durumda RL nasıl yapılabilir? Tartışma: Dataset kalitesi offline RL performansını neden belirler?
13	Taklit Öğrenme ve Tersine Pekiştirmeli Öğrenme	Kavramsal: Behavioral cloning, expert demonstrations ve inverse RL kavramlarını inceleyin. Karşılaştırma: Supervised learning ile imitation learning arasındaki farkı düşünün. Hazırlık: Expert demonstration'lardan doğrudan policy öğrenmenin neden uzun vadede problem yaratabileceğini (covariate shift/distribution shift) inceleyin. Okuma: DAgger makalesi (Ross et al. 2011) ve Maximum Entropy IRL üzerine temel kaynak (Ziebart et al. 2008). Soru: IRL neden doğrudan policy yerine reward function öğrenmeye çalışır?
14	İnsan Geri Bildirimi ile Pekiştirmeli Öğrenme	Ön bilgi: Policy optimization ve reward function kavramlarını tekrar edin. Kavramsal: Human preference, preference data, reward model ve RLHF pipeline'ını inceleyin. Okuma: Christiano et al. 2017 (Deep RL from Human Preferences) ve Ouyang et al. 2022 (InstructGPT). Hazırlık: İnsanların doğrudan reward fonksiyonu yazmasının neden zor olabileceğini düşünün. Tartışma: Reward model yanlışsa (reward hacking) RL agent nasıl davranabilir?

15	Çok Ajanlı Pekiştirmeli Öğrenme-Aktarımlı Pekiştirmeli Öğrenme -Çok Görevli Pekiştirmeli Öğrenme	Kavramsal: Single-agent ve multi-agent RL farkını inceleyin. Cooperative vs competitive environments, centralized training/decentralized execution kavramlarına bakın. Transfer: Domain/task transfer ve generalization kavramlarını inceleyin. Hazırlık: Birden fazla agent olduğunda environment neden non-stationary hale gelir? Tartışma: Öğrenilmiş bir policy başka bir göreve nasıl aktarılabilir?
16	Final	

Değerlendirme Sistemi

Etkinlikler	Sayı	Katkı Payı
Devam/Katılım		
Laboratuar		
Uygulama		
Arazi Çalışması		
Derse Özgü Staj		
Küçük Sınavlar/Stüdyo Kritiği		
Ödev		
Sunum/Jüri		
Projeler		
Seminer/Workshop		
Ara Sınavlar		
Final		
Dönem İçi Çalışmaların Başarı Notuna Katkısı		0
Final Sınavının Başarı Notuna Katkısı		
TOPLAM		0

AKTS İşyükü Tablosu

Etkinlikler	Sayı	Süresi (Saat)	Toplam İşyükü
Ders Saati			
Laboratuar			
Uygulama			
Arazi Çalışması			
Sınıf Dışı Ders Çalışması			
Derse Özgü Staj			
Ödev			
Küçük Sınavlar/Stüdyo Kritiği			

Projeler			
Sunum / Seminer			
Ara Sınavlar (Sınav Süresi + Sınav Hazırlık Süresi)			
Final (Sınav Süresi + Sınav Hazırlık Süresi)			
		Toplam İşyükü	0
		Toplam İşyükü / 30(s)	0.00
		AKTS Kredisi	0

Diğer Notlar	Yok