

# Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme (Reinforcement Learning, RL), bir ajanın (agent) belirli bir ortamda (environment) belirli bir görevi yerine getirirken, deneyimlerinden öğrenerek en iyi stratejiyi (policy) geliştirmesini amaçlayan bir makine öğrenmesi yöntemidir. Ajan, ortamla etkileşime girer ve her eylemi için bir ödül (reward) veya ceza (penalty) alır. Bu ödüller ve cezalar aracılığıyla ajan, uzun vadede en yüksek toplam ödülü elde edecek eylemleri öğrenir.

## Temel Kavramlar

- Ajan (Agent):**
  - Karar veren varlık.
  - Ortamda belirli bir görevi yerine getirmek için eylemler yapar.
- Ortam (Environment):**
  - Ajanın etkileşimde bulunduğu dış dünya.
  - Ajanın eylemlerine tepki verir ve ödüller veya cezalar sağlar.
- Durum (State):**
  - Ortamın belirli bir anındaki durumu.
  - Ajanın karar verirken göz önünde bulundurduğu bilgi.
- Eylem (Action):**
  - Ajanın bulunduğu durumda gerçekleştirebileceği hareket veya karar.
- Ödül (Reward):**
  - Ajanın yaptığı eylemin sonucunda aldığı geri bildirim.
  - Amaç, mümkün olan en yüksek toplam ödülü elde etmektir.
- Politika (Policy):**
  - Ajanın her durumda hangi eylemi seçeceğini belirleyen strateji.
  - Politika, durumlardan eylemlere bir eşlemedir.
- Değer Fonksiyonu (Value Function):**
  - Belirli bir durumun veya durum-eylem çiftinin ne kadar iyi olduğunu gösterir.
  - Uzun vadeli ödül beklentisini temsil eder.

## Pekiştirmeli Öğrenmenin Temel Bileşenleri

- Öğrenme Modelleri:**
  - Model-tabanlı (Model-based):** Ortamın dinamiklerini modellemeye çalışır. Gelecek durum ve ödüller tahmin etmek için kullanılır.
  - Model-siz (Model-free):** Ortamın dinamikleri hakkında bilgi sahibi olmadan, yalnızca deneyimlere dayanarak öğrenir.
- Değer Temelli (Value-based) Yöntemler:**
  - Q-öğrenme (Q-learning):** Eylem-değer fonksiyonu (Q-fonksiyonu) öğrenilir. Her durum-eylem çiftine bir değer atanır ve bu değerler en iyi eylemi belirlemek için kullanılır.
  - Durum-değer fonksiyonu (V-fonksiyonu):** Her duruma bir değer atanır ve bu değerler politika geliştirmek için kullanılır.
- Politika Temelli (Policy-based) Yöntemler:**
  - Doğrudan en iyi politikayı öğrenmeyi amaçlar.
  - Politika Gradyan Yöntemleri:** Politikanın parametrelerini optimize ederek en iyi stratejiyi bulur.
- Hibrit Yöntemler:**

- **Actor-Critic:** Hem deęer fonksiyonunu hem de politikayı aynı anda öğrenir. "Actor" en iyi eylemi seçerken, "Critic" bu eylemin ne kadar iyi olduğunu deęerlendirir.

## Ödöl ve Ceza Yapısı

1. **Anlık Ödöl (Immediate Reward):**
  - Ajanın yaptığı eylemin sonucunda anında aldığı ödöl.
  - Örneęin, bir oyun oynarken her doęru hamlede kazanılan puan.
2. **Gecikmeli Ödöl (Delayed Reward):**
  - Ajanın yaptığı eylemin sonucunda gelecekte aldığı ödöl.
  - Örneęin, bir satranç oyununda hamlelerin sonucunda oyunu kazanmak.
3. **Toplam Ödöl (Cumulative Reward):**
  - Ajanın belirli bir görevi tamamladığında elde ettiği toplam ödöl.
  - Amaç, toplam ödölü maksimize etmektir.

## Algoritmalar

1. **Q-öğrenme (Q-learning):**
  - Model-siz deęer temelli bir öğrenme algoritmasıdır.
  - Q-fonksiyonunu öğrenir ve bu fonksiyon her durum-eylem çiftine bir deęer atar.
  - Ajan, Q-fonksiyonunu kullanarak en iyi eylemi belirler.
2. **SARSA (State-Action-Reward-State-Action):**
  - Q-öğrenmeye benzer ancak ajanın mevcut politikayı takip etmesini sağlar.
  - Bir durumdan başlayıp bir eylem yaparak ödöl alır ve yeni duruma geçer. Sonraki eylemi de göz önünde bulundurarak öğrenme işlemini yapar.
3. **Derin Q-Aęları (Deep Q-Networks, DQN):**
  - Derin öğrenme ve Q-öğrenmeyi birleştirir.
  - Karmaşık ortamlarda, büyük durum-uzaylarında Q-deęerlerini öğrenmek için derin sinir aęları kullanır.
4. **Politika Gradyan Yöntemleri:**
  - Doğrudan politikayı optimize eder.
  - Politikanın parametrelerini, ödöl fonksiyonunu maksimize edecek şekilde günceller.
5. **Proksimal Politika Optimizasyonu (Proximal Policy Optimization, PPO):**
  - Politika gradyan yöntemlerinin stabilitesini artırmak için kullanılır.
  - Eylemlerin büyük deęişimlerini engelleyerek daha güvenilir öğrenme sağlar.

## Uygulama Alanları

1. **Oyunlar:**
  - Atari oyunlarından Go'ya kadar birçok oyunda RL kullanılmıştır.
  - Örneęin, Google DeepMind'ın AlphaGo'su, RL kullanarak insan oyuncularını yenmiştir.
2. **Robotik:**
  - Robotların karmaşık görevleri öğrenmesi ve gerçekleştirmesi için kullanılır.
  - Örneęin, bir robotun belirli bir yolda engelleri aşarak hedefe ulaşması.
3. **Özerk Sistemler:**
  - Otonom araçların ve dronların karar verme süreçlerinde kullanılır.

- Trafikte güvenli ve etkili bir şekilde hareket etmelerini sağlar.
- 4. **Finans:**
  - Hisse senedi ticareti, portföy optimizasyonu gibi alanlarda kullanılır.
  - Finansal piyasalarda stratejik kararlar almak için RL algoritmaları kullanılır.
- 5. **Sağlık:**
  - Tedavi planlarının optimize edilmesi ve hastalık teşhisinde kullanılır.
  - Kişiselleştirilmiş tıbbi tedaviler geliştirmek için kullanılabilir.

## **Sonuç**

Pekiştirmeli öğrenme, bir ajanın belirli bir ortamda etkileşimde bulunarak ve ödüller aracılığıyla en iyi stratejiyi öğrenmesini sağlayan güçlü bir öğrenme yöntemidir. Değer temelli, politika temelli ve hibrit yöntemlerle çeşitli algoritmalar sunan RL, birçok endüstride ve araştırma alanında devrim yaratmış ve yaratmaya devam etmektedir. Karmaşık problemleri çözme yeteneği sayesinde, RL'nin önemi ve uygulama alanları giderek artmaktadır.