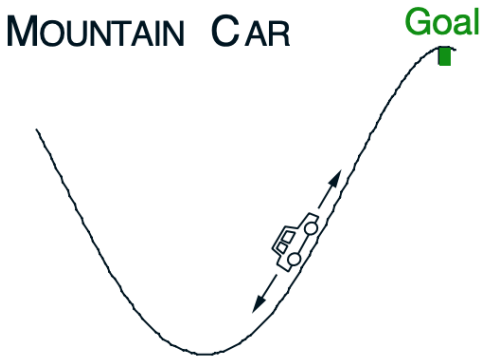
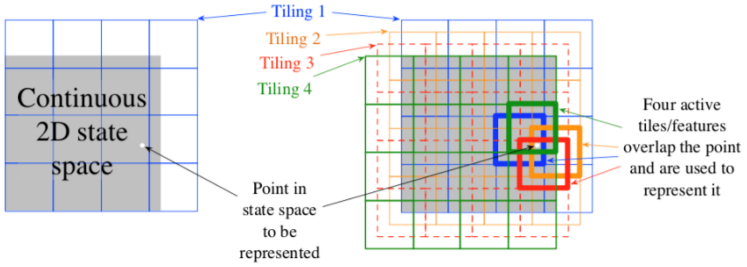
تمرین پنجم – یادگیری تعاملی
نرجس نورزاد – امین تبریزیان
یادگیری در فضای پیوسته

سوال ۱) گسسته سازی با روش Tile Coding

فرض کنید ماشینی در اختیار داریم که باید از دره‌ای به بالا برود. ماشین قدرت کمی دارد بنابراین یادگیر باید طوری به جلو و عقب حرکت کند که گشتاور لازم برای بالا رفتن ماشین را فراهم سازد. در هر زمان، یادگیر از محیط سرعت فعلی (بین ۰.۰۷ و ۰.۰۷-) و مکان فعلی خود (بین ۱.۲- و ۰.۵) را دریافت می‌کند.



از آنجایی که فضای حالت این محیط پیوسته است، تعداد state هایی که یادگیر می‌تواند در آن‌ها قرار بگیرد بی‌شمار است. استفاده از روش های function approximation می‌تواند به یادگیر برای یافتن سیاست بهینه در این نوع از محیط کمک کند. در این سوال قصد داریم از روش tile coding استفاده کنیم. به علاوه، عامل از الگوریتم SARSA برای پیدا کردن سیاست بهینه استفاده می‌کند.



برای آشنایی با الگوریتم tile coding و روش استفاده ی آن، به این لینک مراجعه کنید.

بخشی از کد کلس function approximation لازم و همچنین کتابخانه‌هایی که برای این سوال نیاز دارید در notebook ای که همراه تمرین آپلود شده، در اختیار شما قرار گرفته‌است.

برای ۳ حالت tile coding زیر، نمودار step per episodes / episode را (هر سه را در یک نمودار) رسم کنید و پس از بررسی نتایج بهترین حالت coding برای محیط ارائه شده را تعیین کنید.

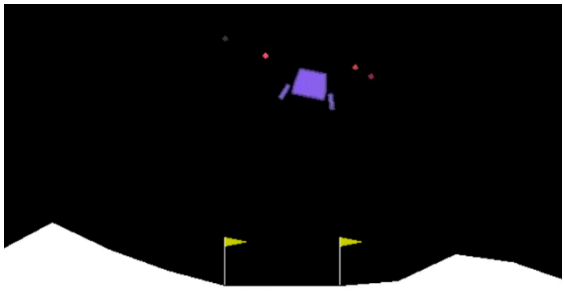
۱. num_tiles: 16, num_tilings: 2, iht_size: 4096

۲. num_tiles: 4, num_tilings: 32, iht_size: 4096

۳. num_tiles: 8, num_tilings: 8, iht_size: 4096

سوال ۲) کنترل فرود فضاپیما

در این سوال قصد داریم با استفاده از روش های یادگیری تقویتی یک یادگیر توسعه دهیم تا بتواند فرود یک فضاپیما را به صورت خودکار کنترل کند. محیطی که برای شبیه‌سازی این سوال استفاده می‌شود LunarLander-v2 از کتابخانه ی gym است.



این محیط متغیر های حالت پیوسته ی زیر را دارد:

$$state \rightarrow \begin{cases} x \\ y \\ v_x \\ v_y \\ \theta \\ v_\theta \\ left\ leg\ touched\ the\ ground \\ right\ leg\ touched\ the\ ground \end{cases}$$

که متغیر های حالت ۱ تا ۶ به ترتیب مربوط به مکان x, y، سرعت در جهت x, y، زاویه در فضا و سرعت زاویه‌ای فضاپیما هستند. دو متغیر حالت پایانی که از نوع Boolean می‌باشند درصورت تماس قسمت چپ یا راست فضاپیما با زمین مقدار ۱ خواهند داشت.

فضای حرکت یادگیر نیز شامل چهار تصمیم گسسته ی عدم حرکت، حرکت به سمت راست، چپ و بالا است.

پاداش یادگیر نیز به‌صورت زیر محسابه می‌شود:

$$R(s) = -100 \times (d_t - d_{t-1}) - 100 \times (v_t - v_{t-1}) - 100 \times (\omega_t - \omega_{t-1}) + landed(s_t)$$

که در آن d فاصله فضاپیما با محل فرود، v سرعت فضاپیما، ω سرعت زاویه‌ای فضاپیما و $landed(s_t)$ تابع پاداش فرود است. این تابع پاداش با توجه به کیفیت فرود فضاپیما پاداش متناسب را تعیین می‌کند.

الف) یک یادگیر بر مبنای الگوریتم deep Q-network توسعه دهید. الگوریتم توسعه داده شده باید شامل دو شبکه ی عصبی برای محاسبه و به‌روز رسانی تخمین‌گر action-value ها، یک بافر برای ذخیره ی تجربه های یادگیر (experience replay) و سیاست $\epsilon - greedy$ باشد. نمودار تابع پاداش یادگیر در طول زمان یادگیری را رسم کنید. همچنین اثر اندازه‌ی بافر بر روی کیفیت یادگیری را بررسی کنید.

ب) *امتیازی:* با جست و جو در مقالات یک یادگیر برمبنای یکی از روش‌های Actor-Critic توسعه دهید و عملکرد آن را با قسمت الف مقایسه کنید.

در هر قسمت شبکه های عصبی مربوط به یادگیر با بهترین عملکرد را ذخیره کنید و همچنین فیلم کوتاهی از عملکرد یادگیر بهینه تهیه کنید.

نکات تکمیلی:

- سعی کنید از پاسخ های روشن در گزارش خود استفاده کنید و اگر پیش فرضی در حل سوال در ذهن خود دارید، حتما در گزارش خود آن را ذکر نمایید.

- حجم گزارش شما به هیچ وجه معیار نمره دهی نیست، پس لطفا در حد نیاز توضیح دهید.

- از نمودارهای واضح در گزارش خود استفاده کنید، نمودارهایتان حتما دارای لیبل واضح روی هر محور و توضیح مناسب باشد.

- لطفا در گزارش و کدهای خود از تمرین دیگران استفاده نکنید. مشورت و هم‌فکری در مورد سوال‌ها اشکالی ندارد اما اگر شباهت بیش از اندازه در تمرین ها دیده شود منجر به صفر شدن نمره خواهد شد.

- تمام فایل‌ها را در قالب یک فایل zip در سایت درس بارگذاری کنید.

- حتما فرمت گزارش که در سایت درس قرار داده شده است را رعایت نمایید.

- در صورت وجود هر نوع سوال در رابطه با این سری تمرین می‌توانید از طریق بخش پرسش و پاسخ سایت ایلرن و هم‌چنین ایمیل‌های زیر با دستیاران آموزشی در ارتباط باشید. از آنجایی که معمولا سوالات به‌وجود آمده برای شما برای سایر دوستانتان نیز وجود دارد توصیه می‌شود تا حد امکان سوالات خود را در فروم مطرح کنید.

- نرجس نورزاد *njnoorzad@gmail.com* (سوال ۱)

- امین تبریزیان *amin.tabrizian@ut.ac.ir* (سوال ۲)

شاد و سلامت باشید :)