

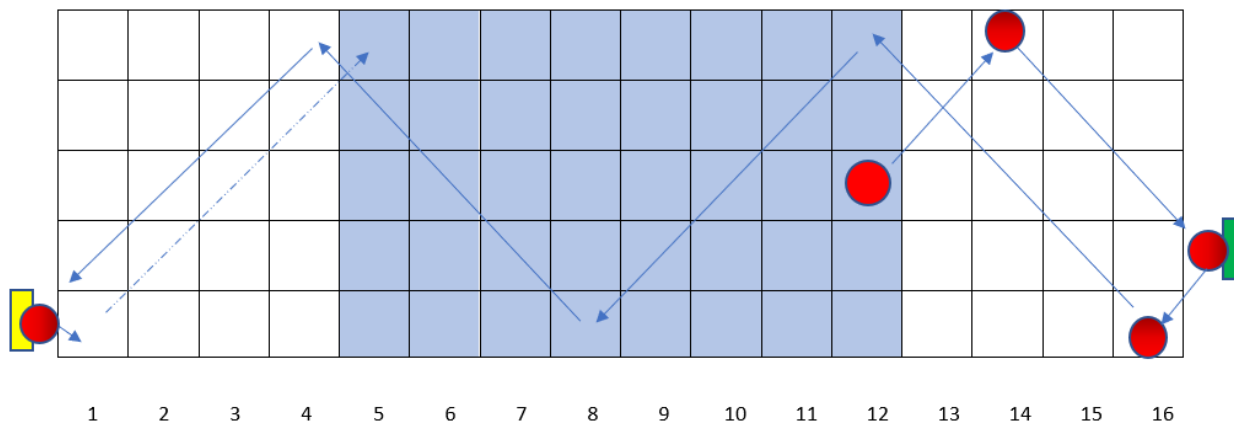
## پروژه نهایی درس یادگیری تقویتی

### مدرس : دکتر خواسته

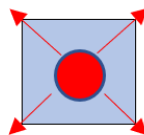
برای پروژه نهایی درس لطفا به یکی از سوالات زیر پاسخ دهید.

### سوال یک

در یک بازی به صورت زیر که مشابه ایرهاکی است دو بازیکن تلاش میکنند توپ از آن‌ها رد نشود.



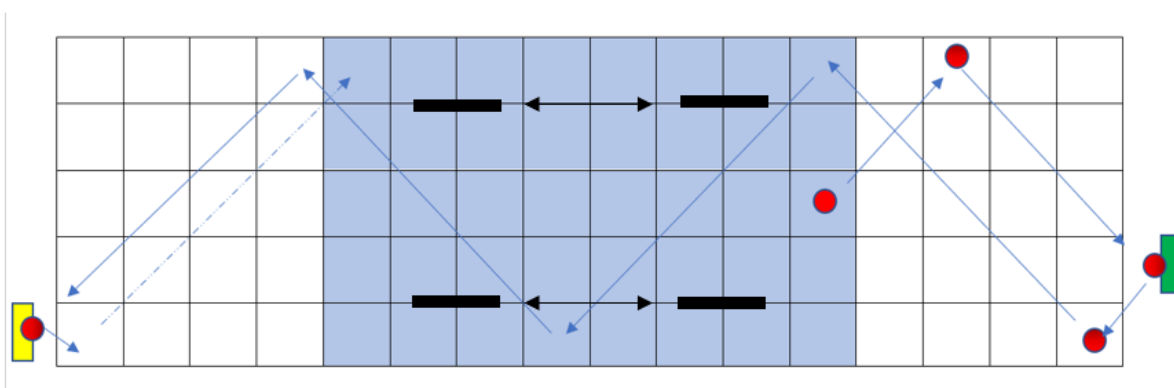
در این بازی اگر بازیکن نتواند توپ را برگرداند یک امتیاز از دست می‌دهد. ایجنت را به گونه‌ای آموزش دهید که بتواند توپ را برگرداند و بازی را ببرد. در ابتدای بازی توپ در یکی از خانه‌های آبی رنگ و در یکی از جهت‌های مشخص شده در شکل زیر شروع به حرکت می‌کند.



**راهنمایی:** ایجنت را با بازی با خودش train کنید. در این بازی می‌توانید فرض کنید بازیکن زرد رنگ در ستون صفرام و بازیکن سبز رنگ در ستون 17 قرار دارد.

موارد زیر مطلوب است:

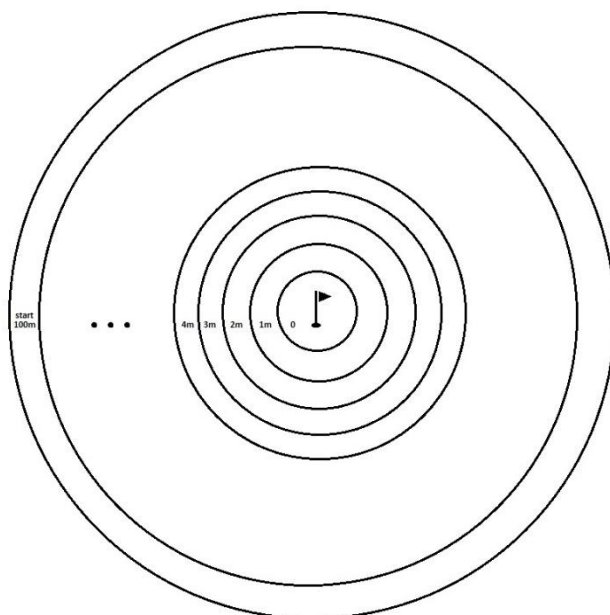
1. مسئله‌ی طرح شده را به روش Q-learning حل کنید، پالیسی بهینه و تابع ارزش state-action را برای هر استیت به دست آوردید.
2. نحوه‌ی انتخاب پاداش، حالات و اکشن‌ها را شرح دهید؟
3. مورد یک را برای این بازی در حالتی که دو مانع متحرک به صورت زیر داشته باشد تکرار کنید. این موانع به این دلیل به بازی اضافه شده اند که بدون حضور آنها گاهی توپ در یک چرخه رفت و برگشتی ثابت گیر می‌کند. سرعت موانع را به گونه ای تنظیم کنید که از این اتفاق در صورت وقوع جلوگیری کند.







4. برای این بازی در نهایت بعد از train ، امکان بازی عامل با انسان را فراهم نمایید. این کار مستلزم پیاده سازی یک گرافیک ساده برای بازی است. (اختیاری)

## سوال دو

زمین گلفی را مانند شکل زیر در نظر بگیرید که در آن شروع بازی از فاصله‌ی صد متری هدف می‌باشد.



به ازای هر ضربه می‌توان از بین چهار چوب Woods، Irons، Hybrids و Putter یک چوب را برای ضربه زدن انتخاب کرد. این چوب‌ها به ترتیب از بی‌دقت‌ترین و پر قدرت‌ترین تا دقیق‌ترین و ضعیف‌ترین مرتب شده‌اند. هر کدام از این چوب‌ها شامل ضریب دقت (precision) و فاصله‌ی پیش‌فرض (default distance) می‌باشند که فاصله‌ی طی شده توسط توپ پس از ضربه با چوب موردنظر را مشخص می‌کند. ضریب دقت و فاصله‌ی پیش‌فرض هر چوب در جدول زیر مشخص شده است.

	Club Type	Default Distance	Precision
	Woods	18 meter	$\mathcal{N}(1, 0.25)$
	Irons	12 meter	$\mathcal{N}(1, 0.15)$
	Hybrids	6 meter	$\mathcal{N}(1, 0.05)$
	Putter	3 meter	1

شدت ضربه (power) با چوب نیز در فاصله‌ای که توپ طی می‌کند تاثیر گذار است که مقدار آن می‌تواند از صفر تا یک با دقت 0.1 (یعنی 10 مقدار) باشد. شدت باد نیز به صورت نرمال می‌تواند در فاصله‌ی طی شده توسط

توپ با مقدار  $\mathcal{N}(0, 3)$  تاثیر گذار باشد. در نهایت رابطه‌ی مسافت طی شده توسط توپ پس از ضربه با چوب  $\hat{l}$  به صورت زیر محاسبه می‌شود:

$$d = [power * default\ distance_i * precision_i] + round((1 - precision_i) * wind_{disturbance})$$

موارد زیر مطلوب است:

1. مسئله‌ی طرح شده را به دو روش SARSA و Q-learning حل کنید، پالیسی بهینه و تابع ارزش state-action را به ازای هر دو روش به دست آورید. (برای هر دو حالت از  $\epsilon - soft\ policy$  با مقدار یکسانی از  $\epsilon$  و  $\alpha$  استفاده کنید).
2. نحوه‌ی انتخاب پاداش، حالات و اکشن‌ها را شرح دهید؟
3. تفاوت عملکرد این دو روش را شرح داده و مشخص کنید که با توجه به نتایج به دست آمده، به نظر شما کدام روش عملکرد بهتری دارد؟
4. شرایطی را در نظر بگیرید که بخش‌هایی از زمین شامل دریاچه بوده و با افتادن توپ در آب بازی به اتمام می‌رسد. این موضوع چه تغییراتی در شرایط مسئله و رویه‌ی اجرای هر دو الگوریتم ایجاد می‌کند؟

توجه کنید که تمام ضربات به سمت هدف بوده و زاویه‌ی ضربه تأثیری در این مسئله و حل آن ندارد. همچنین فاصله‌ی توپ تا هدف در تمامی مراحل اجرا به صورت گسسته می‌باشد.