

بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران

بهار ۱۴۰۰

تحويل: دوشنبه ۱۷ خرداد

تمرین سری یازدهم

مبانی یادگیری عمیق

۱. سیاست (policy) و ارزش (value) در یادگیری تقویتی را تعریف کنید. سپس روش‌های مبتنی بر سیاست و مبتنی بر ارزش را به طور خلاصه توضیح دهید و تفاوت آن‌ها را بیان کنید.

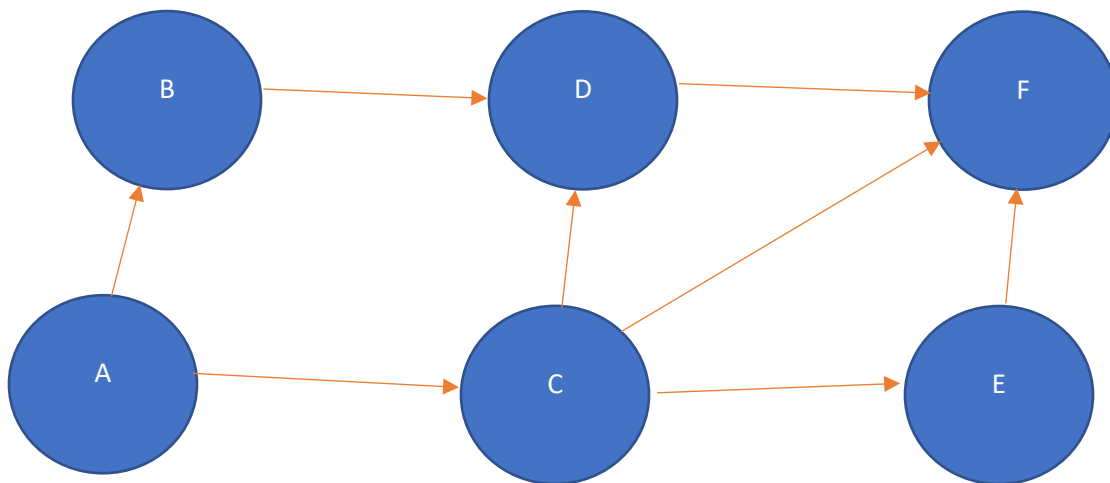
۲. گراف زیر چند state و action‌های مجاز بین آن‌ها را نشان داده است (در یک اپیزود گذشتن از هر state تنها یک بار مجاز است). پاداش‌های دریافتی از state‌های A, B, C, D, E و F (که state پایانی (Terminal) است) به ترتیب عبارتند از: 0, 1, 1-2, 4 و 5 (پارامتر گاما را ۱ در نظر بگیرید)

الف) فرض کنید اگر بر اساس سیاست ۱ پیش برویم، مسیر ACEF انتخاب شود. بازده (return) این مسیر را به دست آورید.

ب) فرض کنید اگر بر اساس سیاست ۲ پیش برویم، مسیر ABDF انتخاب شود. بازده (return) این مسیر را به دست آورید.

ج) بازده کدام یک از سیاست‌های ذکر شده در الف یا ب بهتر بود؟ اگر سیاست با بازده بهتری برای انتخاب مسیر وجود دارد آن را بیان کنید و بازده آن را محاسبه کنید.

د) موارد الف تا ج را با در نظر گرفتن پارامتر گاما برابر با ۰.۵ محاسبه کنید.



۳. در این تمرین قصد این است که یک شبکه Siamese برای یافتن embedding‌های تفکیک‌کننده برای اعداد موجود در مجموعه داده MNIST طراحی شود. در [نوت‌بوک](#) تهیه‌شده پیاده‌سازی‌های لازم را انجام دهید و خروجی خود را که embedding‌ها را با استفاده از PCA در فضای دو بعدی نمایش می‌دهد، تحلیل کنید (از PCA استفاده می‌شود تا ابعاد بردار ورودی را به ۲ کاهش دهیم تا در صفحه قابل نمایش باشد و این کار را هم برای embedding‌های استخراج شده و هم برای خود تصاویر خام انجام می‌دهیم). همچنین، در رابطه با محتوای هر cell که در ابتدای آن عبارت `describe this cell` # نوشته شده است هم به طور دقیق توضیح دهید.

نکات تکمیلی

- (۱) لطفاً پاسخ سوالات (تئوری و توضیحات پیاده‌سازی) را به طور گویا و به زبان فارسی و در صورت امکان تایپ همراه با سورس کدهای نوشته شده، در یک فایل فشرده شده به شکل HW11_YourStudentID.zip قرار داده و بارگذاری نمایید.
- (۲) منابع استفاده شده را به طور دقیق ذکر کنید.
- (۳) برای سهولت در پیاده‌سازی‌ها و منابع بیشتر، زبان پایتون پیشنهاد می‌شود. لطفاً کدهای مربوطه را به طور جداگانه در فرمت ipynb ارسال نمایید.
- (۴) ارزیابی تمرین‌ها براساس صحیح بودن راه حل‌ها، گزارش مناسب، بهینه بودن کدها و کپی نبودن می‌باشد.
- (۵) در مجموع تمام تمرین‌ها، تنها ۷۲ ساعت تاخیر در ارسال پاسخ‌ها مجاز است اما پس از آن به صورت خطی از نمره شما کسر خواهد شد (معادل با روزی ۵۰ درصد).
- (۶) تمرین‌ها باید به صورت انفرادی انجام شوند و حل گروهی تمرین مجاز نیست.
- (۷) در رابطه با پرسش و پاسخ در رابطه با تمرین‌ها می‌توانید در گروه مربوطه مطرح کنید.

موفق باشید.