یادگیری عمیق

نیم‌سال دوم ۹۹-۹۸

دانشکده‌ی مهندسی کامپیوتر

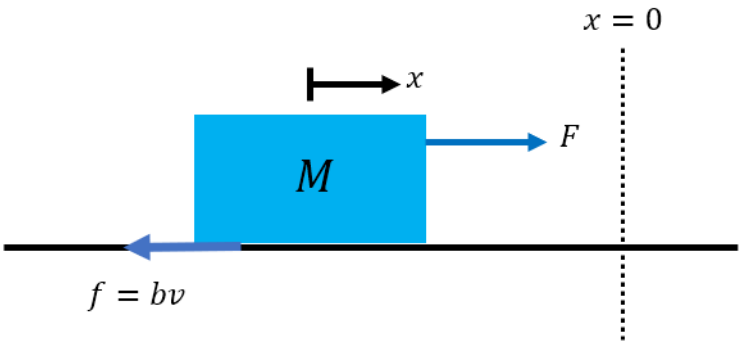
مدرس: دکتر سلیمانی

تمرین سری ششم | یادگیری تقویتی | موعد تحویل: ۲۷ خرداد ۹۹

نمره ۲۰

مسئله‌ی ۱. Policy Gradient (۲۰ نمره)

در این سوال به پیاده سازی یک کنترلر با استفاده از الگوریتم Policy Gradient خواهیم پرداخت.در این مسئله یک جسم با جرم M بر روی یک سطح با اصطکاک ویسکوز حرکت می کند.



که متغیر x مکان جسم و v سرعت جسم است. معادله حرکت به صورت زیر است:

$$M\ddot{x} + b\dot{x} = F \qquad (۱)$$

هدف مسئله طراحی کنترلری است که بتواند از طریق نیروی F این جرم را در نقطه $x = 0$ با سرعت صفر در کمترین زمان ممکن متوقف کند. برای سادگی مسئله و کوانتیزه شدن، نیروی F را به صورت نرمالیزه شده نسبت به جرم جسم به دو صورت $F = M$ و $F = -M$ در نظر می گیریم. با در نظر گرفتن ضریب $k = \frac{b}{M}$ و ورودی نیرو $a = \frac{F}{M}$ و در نظر گرفتن مکان جسم به عنوان حالت اول $(s_۱)$ و سرعت جسم به عنوان حالت دوم $(s_۲)$ معادلات زمان گسسته دینامیک سیستم در زمان t با زمان نمونه برداری T به صورت زیر است:

$$\begin{aligned} s_۱(t+۱) &= s_۱(t) + T(s_۲(t)) \\ s_۲(t+۱) &= s_۲(t) + T(-k.s_۲(t) + a) \end{aligned} \qquad (۲)$$

در واقع $Agent$ شما که یک شبکه عصبی است که با گرفتن دو ورودی مکان و سرعت جسم خروجی $+1$ یا -1 را بدهد. شما باید کدی که در اختیارتان قرار گرفته را کامل کنید. یک تابع $reward$ مناسب برای مسئله انتخاب کنید. جواب این مسئله که از طریق تئوری کلاسیک کنترل بهینه به دست آمده در کد قرار دارد و دقت خروجی شبکه عصبی شما با این تابع سنجیده می شود. که باید بالای ۸۰ درصد دقت داشته باشد. نمودار متوسط $reward$ در هر $epoch$ را در حین یادگیری رسم کنید.

موفق باشید