

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

تمرین سری چهارم یادگیری ماشین

دانشکده مهندسی کامپیوتر

استاد درس: دکتر ناظر فرد

خرداد ۱۴۰۰

- تمامی مستندات شامل گزارش به همراه کدهای خود را در یک فایل فشرده با فرمت zip ذخیره کرده و با عنوان stdNum_HW2.zip که stdNum شماره دانشجویی شما است در سامانه بارگزاری کنید.
- سوالات ستاره‌دار(*) نمره اضافی داشته و انجام آن‌ها اجباری نمی‌باشد.
- مهلت انجام تمرین تا ساعت ۲۳:۵۵ روز چهارشنبه ۲ تیر می‌باشد.

سوال‌های تشریحی (یادگیری بدون نظارت، خوشه‌بندی، یادگیری تقویتی)

سوال ۱) به سوالات زیر در مورد الگوریتم‌های خوشه‌بندی پاسخ دهید.

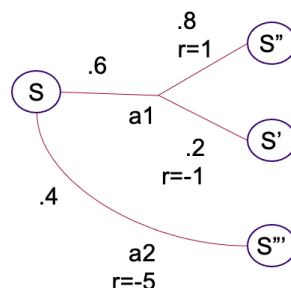
- الف)** معیارهای مختلف تعیین مرکز برای الگوریتم خوشه‌بندی K-Means را بیان کنید و مزایا و معایب هر کدام را توضیح دهید.
- ب)** معیارهای مختلف فاصله برای الگوریتم K-Means را بیان کنید و مزایا و معایب هر کدام را بیان کنید.
- ج)** تاثیر نرمال‌سازی داده‌ها را پیش از اجرای الگوریتم K-Means بررسی کنید.
- د)** معیارهای مختلف فاصله (از نظر انتخاب نمونه‌ها) در الگوریتم خوشه‌بندی سلسله مراتبی را بیان کنید و آن‌ها را از نظر پیچیدگی محاسبات مقایسه کنید. همچنین در مورد مزایا و معایب هر کدام از این معیارهای توضیح دهید.
- ه)** الگوریتم خوشه‌بندی سلسله مراتبی بالا-به-پایین و پایین-به-بالا را مقایسه کنید و برای کاربرد هر کدام مثالی ارائه کنید.
- و)** روشی برای تعیین هایپرپارامترها در الگوریتم خوشه‌بندی DBScan ارائه کنید. مزایا و معایب این الگوریتم خوشه‌بندی را نسبت به سایر روش‌ها بررسی کنید.

سوال ۲) به سوالات زیر در مورد یادگیری تقویتی پاسخ دهید.

- الف)** یادگیری تقویتی و یادگیری بانظارت از دو جنبه با هم متفاوت می‌باشد. آن دو جنبه را توضیح دهید.
- ب)** تفاوت دو الگوریتم Policy Iteration و Value Iteration با یکدیگر چیست؟

سوال ۳) با توجه به شکل زیر به سوالات پاسخ دهید.

- الف)** تابع مقدار^۱ را بدست آورید.
- ب)** چرا گاما در محاسبات تاثیر نداشت؟



^۱ Value Function

- کدهای خود را به زبان پایتون و ترجیحا در محیط jupyter پیاده‌سازی کنید. می‌توانید تحلیل خودتان را به عنوان سلول‌های متنی در همان محیط ارائه کنید.
- نظم در نوشتن گزارش و کدها می‌تواند به کسب نمره‌ی بهتر به شما کمک کند. برنامه نوشته شده خوانا و کامنت گذاری مناسب داشته باشد.
- در پیاده‌سازی بخش‌های مختلف، امکان استفاده از کتابخانه‌های آماده مرتبط با الگوریتم‌های یادگیری ماشین را به طور کلی ندارید. **موارد مجاز در صورت سوال بخش‌ها ذکر شده است.**
- برای نمایش نمودارها و عملیات ماتریسی می‌توانید از کتابخانه‌های numpy و matplotlib استفاده کنید.
- همچنین برای خواندن داده‌ها به عنوان ورودی می‌توانید از pandas استفاده کنید.
- برای محاسبه معیارهای ارزیابی مانند دقت و ماتریس درهم‌ریختگی می‌توانید از کتابخانه آماده استفاده کنید.
- مطابق قوانین دانشگاه هرگونه کپی‌برداری ممنوع می‌باشد و در صورت مشاهده نمره هر دو طرف صفر در نظر گرفته می‌شود.
- در صورت داشتن سوال می‌توانید با ایمیل تدریس یاران درس در تماس باشید:
hse.khalilian08@gmail.com , hamid.dargahi0072@gmail.com

سوال‌های پیاده‌سازی

مسئله ۱) الگوریتم خوشه‌بندی KMeans را (بدون استفاده از کتابخانه‌های آماده برای این منظور) پیاده‌سازی کرده و عمل خوشه‌بندی را برای مجموعه داده‌ی ۱ question در پوشه تمرین انجام دهید. برای انتخاب تعداد خوشه‌ها عمل خوشه‌بندی را برای Kهای متفاوت انجام دهید و بهترین مقدار را با توجه به معیار تدریس شده در کلاس انتخاب کنید. رسم نمودار مربوطه برای این انتخاب ضروری است.

پس از انتخاب تعداد مناسب خوشه‌ها عمل خوشه‌بندی را انجام داده و مراحل بروز رسانی مرکز خوشه‌ها را در کنار نمونه‌ها رسم کنید. رنگ نمونه‌ها را به گونه‌ای انتخاب کنید که نشان‌دهنده‌ی نتیجه‌ی نهایی خوشه‌بندی باشد. همچنین نمودار SSE را برای حسب تکرارهای مختلف رسم کنید.

انتخاب معیار فاصله اختیاری است. مرکز ابتدایی خوشه‌ها را به صورت تصادفی در نظر بگیرید.

مسئله ۲) الگوریتم خوشه‌بندی DBSCAN را پیاده‌سازی کرده و برای مجموعه داده‌ی ۲ question، عمل خوشه‌بندی را با استفاده از این الگوریتم انجام دهید. داده‌ها را با استفاده از رنگ‌بندی مناسب برای هر خوشه رسم کنید. پس از اجرای خوشه‌بندی معیار purity را برای آن محاسبه کنید.

مسئله ۳) الگوریتم خوشه‌بندی سلسله مراتبی را یک‌بار به صورت بالا-به-پایین و یک‌بار به صورت پایین-به-بالا برای مجموعه داده‌ی ۳ question و با سه معیار مختلف فاصله (از نظر انتخاب نمونه‌ها برای محاسبه‌ی فاصله بین دو خوشه) پیاده‌سازی کنید.

تعداد نهایی خوشه‌ها را ۸ در نظر بگیرید و مراحل مختلف خوشه‌بندی را در حالت بالا-به-پایین از ۲ خوشه به ۸ خوشه و در حالت پایین-به-بالا از ۸ خوشه به ۲ خوشه ترسیم کنید (رسم نمونه‌ها با رنگ‌های متفاوت مربوط به هر خوشه). تاثیر هر یک از معیارهای فاصله را در خوشه‌بندی بررسی کنید .

***مسئله ۴)** این مسئله برای آشنایی شما با الگوریتم‌های یادگیری تقویتی مدرن طراحی شده‌است. از این رو نمره‌ی آن به صورت اختیاری ولی دارای مقداری قابل توجه خواهد بود. یکی از تحقیقاتی که تاثیر به سزایی در رشد مسائل یادگیری تقویتی داشته‌است مقاله‌ی گروه deepmind است که لینک آن در زیر قرار داده شده‌است.

<https://storage.googleapis.com/deepmind-media/dqn/DQNNaturePaper.pdf>

این روش که تحت عنوان مدل‌های Deep Q Network شناخته می‌شود، تحولی بزرگ در یادگیری تقویتی ایجاد کرده است. در این مساله شما باید اطلاعاتی در مورد این روش کسب کنید و عاملی برای بازی در محیط Mountain Car آموزش دهید. تصویر این محیط در شکل ۲ آمده‌است.



هدف آموزش عاملی است که ماشین متحرک در قعر دره را با کمترین هزینه به قله هدایت کند. برای آموزش این عامل می‌توانید مثال زیر را به گونه‌ای تغییر دهید که به جای آموزش عامل برای محیط Cartpole آن را برای محیط MountainCar آموزش دهد .

https://www.tensorflow.org/agents/tutorials/1_dqn_tutorial

در این مثال متغیر `env_name = 'CartPole - v0'` باید به `env_name = 'MountainCar - v0'` تغییر کند. می‌توانید از لینک زیر برای آشنایی بیشتر با یادگیری تقویتی استفاده کنید.

<https://medium.com/ai%C2%B3-theory-practice-business/reinforcement-learning-part-1-a-brief-introduction-a53a849771cf>

همچنین با جستجوی کلید واژه‌ی Deep Q Network و DQN می‌توانید اطلاعات بیشتری در این زمینه کسب کنید. داشتن دانش لازم برای حل مساله و همچنین حل مساله به صورت مجزا در این سوال دارای امتیاز است .