

## تمرین کامپیوتری سوم درس مقدمه‌ای بر یادگیری ماشین - گروه ۱

نیمسال دوم ۱۴۰۳-۱۴۰۲

موعد تحویل: ۱۸ خرداد ۱۴۰۳

آداب انجام تمرینات کامپیوتری

- **همفکری** دانشجویان در انجام تمرینات مانعی ندارد ولی تمرینات باید به صورت مشخص توسط خود تحویل دهنده انجام گرفته باشد. در زمان تصحیح تمرینات شباهت تمرینات بررسی خواهد شد و تمرین‌های با شباهت غیرعادی مورد قبول نیست.
- استفاده از منابع اینترنتی برای گرفتن ایده‌ی حل سوالات مانعی ندارد ولی کپی کردن پاسخ تمرینات از هر منبعی مورد قبول نیست.
- استفاده از کتابخانه‌های آماده در صورتی که در صورت تمرین قید نشده باشد، مانعی ندارد.
- کدهای تحویل شده باید قابلیت اجرای دوباره را داشته باشند. در صورت وجود مشکل در زمان تصحیح (حتی به دلیل خطای تایپی)، قسمت‌های غیرقابل اجرا **مورد قبول نیست**. لطفاً قبل از تحویل تمرین حتماً کد خود را بررسی فرمایید.
- فایل ZIP شامل کدها و خروجی به همراه فایل گزارش (توضیحات در مورد آنچه انجام دادید و سوالات پرسیده شده) در CW آپلود گردد (در صورت استفاده از فرمت ipynb، می‌توانید توضیحات را در همان نوتبوک بنویسید و نیازی به گزارش نیست).
- سوالات و ابهامات در رابطه با این تمرین را می‌توانید در تلگرام با آیدی @mehrdad\_a\_e\_o مطرح کنید.

کاهش بعد و خوشه بندی

تمرین سری سوم

سوال اول :

فایل نوتبوک ML-HW3-Q1 را باز کرده و طبق توضیحات گفته شده داخل فایل ، آن را تکمیل کنید. در بخش اول این تمرین، به پیاده سازی آنالیز مولفه‌های اساسی (PCA) بر روی مجموعه داده MNIST می‌پردازیم. در بخش دوم، یک شبکه Auto-encoder را پیاده سازی خواهید کرد و تفاوت روش اول با این شبکه را بررسی می‌کنید.

مجموعه داده MNIST حاوی تصاویری از ارقام دست نویس است. برای خواندن این دیتاست می‌توانید از کتابخانه‌های مختلف استفاده کنید.

سوال دوم :

در این سوال شما باید با استفاده از تکنیک های خوشه بندی، به تجزیه و تحلیل مجموعه داده ها و گروه بندی گل ها بر اساس شباهت های ویژگی آن ها بپردازید.

مجموعه داده Iris flowers یک دیتاست شامل چند ویژگی از گل های زنبق می باشد که می توان از آن ها برای دسته بندی انواع این گل (۳ انواع آن) استفاده کرد.

۱- دیتاست iris داده شده را بخوانید و پیش پردازش های لازم را انجام دهید.

۲- الگوریتم K-means را از ابتدا پیاده سازی کنید.

۳- خوشه بندی K-means را برای گروه بندی به دسته های k پیاده سازی کنید. با مقادیر مختلف k آزمایش کنید و نتایج خوشه بندی را با استفاده از معیارهایی که در درس معرفی شده ارزیابی کنید.

۴- حال الگوریتم Gaussian Mixture Model (GMM) را از ابتدا پیاده سازی کنید.

۵- با در نظر گرفتن  $k=3$  ، الگوریتم GMM را بر روی دیتاست اعمال کنید و نتیجه اعمال این الگوریتم بر داده ها را برای چند iteration با استفاده از نمودار های مناسب رسم کنید. همچنین نتیجه نهایی را نیز رسم کنید.

۶- در مورد نقاط قوت و محدودیت های هر تکنیک خوشه بندی و مناسب بودن آنها برای مجموعه داده Iris بحث کنید.