

تمرین شماره یک
دکتر زارع

Machine Learning and Data Mining

مهلت تحویل اول
۹۸/۰۷/۲۱ ساعت ۲۳:۰۰
مهلت تحویل دوم
۹۸/۰۷/۲۲ ساعت ۲۳:۰۰

۱. دسته بندی با استفاده از k-Nearest Neighbors

هدف از این تمرین پیاده سازی دسته بند kNN و استفاده از آن جهت پیش بینی داده های ورودی می باشد. دانشجویان می بایست این تمرین را بدون استفاده از هیچ کتابخانه آماده ای به زبان های دلخواه Python یا Matlab پیاده سازی نموده و در موعد مقرر تحویل نمایند.

اطلاعات دادگان این تمرین به شرح ذیل می باشد.

Title: Iris Data Set

Link: <https://archive.ics.uci.edu/ml/datasets/iris>

وظیفه های تعریف شده به شرح ذیل می باشد.

- ۱,۱. تشریح عملکرد الگوریتم kNN
- ۱,۲. رسم توزیع داده های موجود در هر ویژگی و توضیح در مورد مشخصه های هر توزیع نظیر میانگین، واریانس، چولگی
- ۱,۳. تقسیم داده ها به صورت ۸۰-۲۰ برای آموزش و تست
- ۱,۴. انتخاب بهترین مقدار k در بازه ۳ تا ۱۵ و رسم خطای مدل به ازای هر k
- ۱,۵. محاسبه متریک های Precision و Recall برای مدل بدست آمده با استفاده از بهترین k

نمره اضافه

رسم فضای حالت داده ها با استفاده از الگوریتم PCA در فضای دو بعدی (برای این کار می توانید از کتابخانه های آماده استفاده نمایید) و رنگ آمیزی سمپل ها در نمودار بر اساس کلاس آنها

نکات تحویل تمرین

۱. مستندسازی بخشی از نمره را شامل می شود. لذا در مستندسازی دقت و سلیقه کافی داشته باشید.

۲. تاریخ تعیین شده به عنوان "مهلت تحویل اول" تاریخ اصلی بوده و در صورت ارائه تمرین در تاریخ "مهلت تحویل دوم" ۳۰ درصد نمره کسر شده و بررسی تمرین بر مبنای ۷۰ درصد نمره اصلی تمرین انجام خواهد شد.
۳. بعد از مهلت تحویل دوم هیچ نمره ای به تمرین داده نخواهد شد.
۴. هرگونه کپی برداری باعث از دست رفتن کل نمره خواهد شد.
۵. تحویل حضوری متعاقبا اعلام خواهد شد.
۶. ایمیل باید با عنوان HW_PRML_#STDNUMBER_#STDNAME_#HWNUMBER ارسال شود.
۷. تمرین ها را به ایمیل sajjadaghapour@ut.ac.ir ارسال نمایید.

موفق باشید