



مهلت تحویل اول

۲۳:۰۰ ساعت ۹۸/۰۹/۰۳

مهلت تحویل دوم

۲۳:۰۰ ساعت ۹۸/۰۹/۰۴

Machine Learning and Data Mining

تمرین شماره سه

دکتر زارع

دسته بندی با استفاده از Decision Tree

هدف از این تمرین آشنایی با الگوریتم Decision Tree و دسته بندی با استفاده ای آن می باشد. نکات انجام تمرین به شرح ذیل می باشد:

۱. دانشجویان تنها برای پیاده سازی الگوریتم و ارزیابی آن مجاز به استفاده از کتابخانه نمی باشند.
۲. استفاده از امکانات و روش های پیشرفته برنامه نویسی در نمره دانشجویان تاثیر مثبت خواهد داشت. به عنوان مثال استفاده از کتابخانه های مناسب در فراخوانی داده ها، کار با آرایه ها، برنامه نویسی شی گرا و غیره.
۳. بسیاری از نکات مربوط به انجام یک وظیفه یادگیری ماشین در صورت تمرین ارائه نشده ولی از دانشجویان انتظار انجام آن وجود دارد. به عنوان مثال پاک سازی داده ها در صورت وجود داده های غلط یا استفاده درست از دیتاست های Train و Test از این دست موارد می باشد. در صورت عدم انجام این موارد از دانشجویان نمره کسر خواهد شد.

اطلاعات دادگان این تمرین به شرح ذیل می باشد.

Title: Heart Disease Data Set

Link: <https://archive.ics.uci.edu/ml/datasets/heart+Disease>

وظیفه های تعریف شده به شرح ذیل می باشد.

۱. رسم توزیع هر ویژگی و توصیف آن
۲. آموزش مدل بر اساس Decision Tree با فرایند Cross Validation
- a. مدل با استفاده از دو معیار جداسازی Information Gain و Gini Index باید آموزش داده شود
۳. اعمال Pruning روی مدل و مقایسه دقت بدست آمده در مدل با مدل استاندارد بدست آمده در مرحله ۲
۴. نمایش درخت بدست آمده پس از اعمال Pruning در هر قالب نمایشی (گراف یا متن)

نمره اضافه

۱. استفاده از Grid Search جهت بدست آورده بهترین پارامترهای مناسب در مدل
۲. مقایسه دقت بدست آمده از مدل پیاده سازی شده با الگوریتم های آماده در کتابخانه های موجود (به عنوان مثال Scikit-learn در پایتون)

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 / فایل باید با عنوان / نام HW_MLDM_#STDNUMBER_#STDNAME_#HWNUMBER ارسال شود.
۷. تمرین ها باید به آدرس های sajjadaghapour@ut.ac.ir, ah.havvaei@ut.ac.ir, e.sadeqi.n@gmail.com ارسال شود.