



هوش محاسباتی – تکلیف سرک دوم

ددلاین: هجدهم آبان

مدرس: دکتر ملک

نیمسال اول ۱۴۰۰-۱۴۰۱

مسائل:

الگوریتم‌های دسته‌بندی KNN و SVM را بر روی دیتاست train.csv اجرا کنید. لازم به ذکر است در این دیتاست چالش‌هایی از قبیل مدیریت داده‌های ناقص مطرح است که باید به طور موثر توسط دانشجو برطرف شود. برای train کردن مدل، می‌بایست دیتای خود را به دو زیربخش train set و validation set تقسیم کنید. (برای این منظور می‌توانید از کتابخانه‌های آماده استفاده کنید). در نهایت دیتاست test.csv را پیش‌بینی (دسته‌بندی) کنید.

شماتیک دیتاست و تعاریف ستون‌ها: (ستون سبز رنگ، ستون هدف است.)

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

نکات:

- دیتاست‌های `train.csv` و `test.csv` را می‌توانید از [اینجا](#) بارگیری کنید.
- پیاده‌سازی KNN باید به طور کامل توسط دانشجو و بدون استفاده از توابع آماده انجام شود.
- می‌توانید برای پیاده‌سازی SVM از توابع موجود در کتابخانه `sklearn` استفاده کنید.
- در صورت هرگونه **تقلب** نمره‌ی این تمرین برای شما ۰ لحاظ می‌شود
- استفاده از زبان غیر پایتون مجاز **نمی‌باشد**.
- این تمرین، تحویل حضوری ندارد؛ بنابراین نوشتن داکيومنت الزامی و بخشی از نمره می‌باشد. همچنین مستندات خود را در همان محیط نوت بوک و در قالب cell های مجزا با فرمت Markdown بنویسید.
- **لینک عمومی نوت بوک** خود را در سامانه LMS آپلود کنید.
- همچنین فایل‌های خود را فشرده کرده و با فرمت زیر در سامانه آپلود نمایید.

HW2-[Full Name]-[Student ID]

موفق باشید
گروه حل تمرین