

بسم الله الرحمن الرحيم

یادگیری ماشین

تمرین سوم

تهیه شده توسط:

فاطمه ساوینی

میلاد مسرور

(1) از جنس Regression می باشد – بله امکان پذیر است – اولین نکته ای که میتوان اشاره کرد این است که وجود دیتاست کافی خیلی کمک شایانی خواهد کرد. بعد از در دست داشتن دیتاست کافی، برای بهتر کردن عملکرد مدل باید روی دیتاست مورد نظر یکسری روش های پیش پردازش داده پیاده سازی کنیم. از جمله این روش ها میتوان به:

missing value detection , feature selection , outlier and noise removal , feature decorrelation اشاره کرد. برای عملکرد بهتر مدل ساخته شده بهتر است hyper parameter tuning انجام دهیم. همچنین hyper parameter tuning ممکن است از Overfitting و Underfitting جلوگیری کند. برای این نوع مسئله خاص به نظر ما بهتر است به دید time series forecasting نگاه کنیم.

(2) با توجه به شکل مجموعه دادگان ما حدس میزنیم که KNN از همه ضعیف تر عمل کند. اما بر روی این دادگان ما سه روش (Logistic Regression (LR), SVM, KNN در پایتون پیاده سازی کردیم (فایل jupyter notebook پیوست گردیده است) و برای دادگان train و test ، برای هر مدل یک score خروجی گرفتیم. نتایج به صورت زیر میباشد :

**A. LR: train score and test score = 0.48 and 0.5**

**B. KNN: train score and test score = 0.95 and 0.45**

**C. SVM (kernel = rbf): train score and test score = 0.5 and 0.49**

**D. SVM (kernel = linear): train score and test score = 0.5 and 0.5**

مدل KNN بشدت دچار Overfitting شده است. مدل های SVM و LR خیلی شبیه به هم عمل کردند اما بنظر ما SVM با kernel = linear برای این دادگان خیلی مناسب میباشد. ولی با این حال دقت مدل های SVM و LR هم زیاد بالا نمیباشد اما حداقل دچار Overfitting نشده اند.

(3) خیر در مسائل چند کلاسه هم می توان استفاده کرد – برای مسائل چند کلاسه میتوانیم یک به یک با هم کلاس ها را مقایسه کنیم و در نهایت از پارامتر Recall و Precision میانگین بگیریم.

(4) بنظر ما با تغییر Activation Function مدل یادگیری ماشین میتوانیم این امکان را فراهم کنیم. یعنی از یک Activation Function استفاده کنیم که برای داده هایی که بصورت خطی جدا ناپذیر هستند، مناسب باشد. یک نظردیگر هم که ما داریم این است که مثلاً برای مدل SVM برای دادگان خطی بهتر است از کرنل Linear استفاده کنیم و برای دادگانی که به صورت خطی جداناپذیر هستند بهتر است از کرنل rbf استفاده کنیم.