

# تمرین ۵

سینا اسکندری ۹۷۵۲۱۰۵۴

-۲

این داده ها در بازه های زمانی طولانی هستند که باعث به وجود آمدن مشکل ناپدید شدن گرادیان (vanishing gradient) می شود. هم چنین این داده ها بازه ها و پترن های یکسان ندارند که در نتیجه مدل نمیتوان به خوبی آموزش ببیند.

-۵

چون در مسئله تشخیص ناهنجاری داده ها بالانس نیستند و تعداد کلاس های صفر بسیار بیشتر از یک هست معیار accuracy ممکن است غلط انداز باشد چون اگر به طور مثال درصد داده کلاس یک برابر ۵ درصد باشد و مدل ما همیشه خروجی صفر بدهد، مقدار accuracy برابر ۹۵ درصد خواهد بود که در مسائل عادی دقت بالایی محسوب می شود اما در این مسئله مدل بسیار بد عمل کرده است و اگر با توجه به accuracy مدلی را انتخاب می کردیم به مشکل می خوردیم.

مقدار loss نیز نمی تواند خوب باشد چون این متریک برای بهینه سازی است و برای ارزیابی مناسب نمی باشد و همچنین مثل بقیه متریک ها نرمال نیست و نمیتوان با استفاده از آن مقایسه کرد.

باقی سوالات در نوتبوک توضیح داده شده است.

منابع:

<https://www.analyticsvidhya.com/blog/2021/06/5-techniques-to-handle-imbalanced-data-for-a-classification-problem/#:~:text=Imbalanced%20data%20refers%20to%20those,very%20low%20number%20of%20observations.>

<https://datagy.io/python-z-score/>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.iqr.html>