

به نام خدا



دانشکده مهندسی کامپیوتر

درس یادگیری عمیق

دکتر مرضیه داودآبادی

تمرین سری پنجم

دستیاران آموزشی:

صادق جعفری

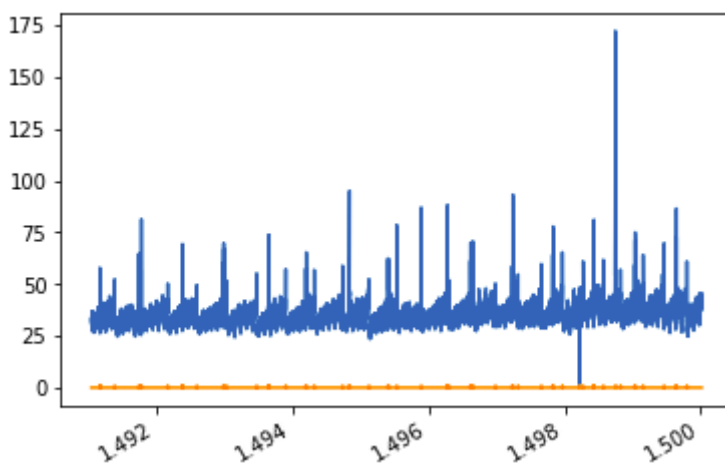
امیرحسین احمدی

نکات تکمیلی

در این تمرین قرار است با استفاده از مدل‌های RNN به تشخیص ناهنجاری‌های موجود در مجموعه داده سری زمانی پرداخته شود. در این مسئله تعدادی داده سری زمانی از دامنه‌های مختلف در اختیار شما قرار داده شده است و وظیفه شما طراحی یک مدل برای تشخیص خودکار ناهنجاری در دامنه‌های مختلف است. منظور از دامنه این است که ممکن است Time step، کران بالا و کران پایین هر سری از داده‌ها متفاوت باشد. برای مثال ممکن است یک مجموعه داده، داده‌هایی بین ۱- تا ۱ و مجموعه‌ای دیگر داده‌هایی بین ۱۰۰۰- تا ۱۰۰۰ داشته باشد.

این تمرین یک مسئله کلاس بندی باینری است، کلاس ۱ نشان‌دهنده این است که داده مورد نظر خراب و ناهنجار است و ۰ نشان‌دهنده این است که داده مورد نظر مشکلی ندارد. توجه: در این مسئله متریک مورد نظر [F1-score](#) است و توجه شما باید روی این باشد که این معیار را افزایش دهید.

۱- ابتدا داده‌ها را Load کنید و آن‌ها را در یک نمودار رسم کنید. (۵ نمره)



عکس 1 یک نمونه از نمودار مورد نظر برای یک مجموعه داده سری زمانی

نمودار آبی نشان دهنده تغییرات سری زمانی و خط نارنجی نشان‌دهنده برچسب هر Time step است.

۲- نمودارها را تحلیل کنید و چالش‌های آموزش این داده‌ها در یک مدل RNN را نام ببرید. (ذکر حداقل ۲ مورد) (۱۰ نمره)

۳- سه مدل RNN (Simple, LSTM, GRU) طراحی کنید و داده‌ها را بدون هیچ پیش‌پردازشی به این مدل‌ها بدهید. آن‌ها را آموزش داده و سپس نتایج را با هم مقایسه کنید. (۱۵ نمره)

۴- چند روش برای پیش‌پردازش داده‌ها ارائه دهید که باعث شود یادگیری آن‌ها توسط مدل‌ها ساده تر شود. (حداقل ۲) (۱۰ نمره)

۵- توضیح دهید چرا معیارهای دیگر نظیر Loss و Accuracy برای ارزیابی مدل‌ها مناسب نیستند. (۱۰ نمره)

۶- در این مسئله بررسی می‌کنیم که آیا میتوان با استفاده از Pre-training نتیجه بهتری نسبت به مدل‌هایی که در مراحل قبل پیاده‌سازی شده است گرفت یا خیر. به آموزش یک مدل بر روی یک تسک یا مجموعه داده و استفاده از پارامترها و وزن‌های آن در مدل دیگری برای یک تسک یا مجموعه داده‌ای متفاوت، Pre-training می‌گویند.

- ابتدا یک تسک [Self-supervised](#) مرتبط با مسئله اصلی تعریف کنید. (راهنمایی: تسک تعریف شده میتواند پیش‌بینی سری زمانی در Time step های بعدی باشد. یا میتوان تعدادی از داده‌ها را حذف کرد و تسک شما این باشد که مدل داده‌های حذف شده را تخمین زده و بازایی کند).
- یک مدل برای تسک تعریف شده طراحی و پیاده‌سازی کرده و آموزش دهید.
- لایه‌های انتهایی مدل را طوری تغییر داده تا مدل متناسب با تسک اصلی شود.
- لایه‌های ابتدایی را فریز کرده و مدل را روی تسک اصلی Fine-tune کنید.

آیا نتیجه بهتری حاصل شد؟ (۳۰ نمره)

۷- داده‌های ارائه شده در این تمرین بسیار نامتعادل‌اند. (تعداد داده‌ها با برچسب ۰ بسیار بیشتر از داده‌های با برچسب ۱ است). یک راهکار برای حل این مشکل ارائه کنید و با اعمال آن عملکرد مدل را بهبود دهید. (۲۰ نمره)

\*\*\* امتیازی \*\*\*

۸- با استفاده از یک [روش آماری](#) سعی کنید این مسئله را حل کنید. آیا با استفاده از روش‌های آماری نتایج بهتری حاصل می‌شود؟ (۲۰ نمره)

## نکات تکمیلی

موفق باشید