



پروژه پنجم

هدف: آشنایی با شبکه خودکدگذار و مقدمه‌ای بر تحلیل سری‌های زمانی با شبکه‌های بازگشتی

کد: کد این فعالیت را به زبان پایتون بنویسید .

گزارش: ملاک اصلی انجام فعالیت، گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf تهیه کنید و در آن برای هر سوال، ورودی، خروجی و توضیحات کامل و جامعی تهیه کنید.

تذکر ۱: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن غیرمجاز است.

تذکر ۲: مجموعه‌های داده مورد استفاده را به جز در مواردی که صریحاً در صورت سوال ذکر شده باشد، حتماً قبل از استفاده بصورت تصادفی به سه بخش آموزش (۷۰ درصد داده‌ها)، آزمون (۲۰ درصد داده‌ها) و اعتبارسنجی (۱۰ درصد داده‌ها) تقسیم نمایید.

راهنمایی: در صورت نیاز میتوانید سوالات خود را در خصوص پروژه از تدریس‌یارهای درس، از طریق ایمیل زیر یا در گروه تلگرامی بپرسید.

Email: ann.ceit.aut@gmail.com

ارسال: فایل‌های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW05.zip تا تاریخ ۱۴۰۱/۱۰/۰۱ ارسال نمایید. شایان ذکر است هر روز تاخیر باعث کسر ۱۰٪ نمره خواهد شد.

بخش اول) شبکه‌های بازگشتی و کاربرد آن‌ها در تحلیل سری‌های زمانی

هدف از تحلیل سری‌های زمانی، استخراج الگو یا پیش‌بینی رفتار یک سری زمانی در طول عمر آن می‌باشد. این مسئله شامل مسائل جزئی‌تر مانند پیش‌بینی، دسته‌بندی، تشخیص ناهنجاری و ... می‌باشد. در حوزه یادگیری عمیق ساختارهای زیادی از جمله شبکه‌های کانولوشنی، شبکه‌های بازگشتی و شبکه‌های ترنسفورمر^۱ برای پردازش داده‌های زمانی معرفی شده‌اند. در این بخش با تمرکز بر شبکه‌های کانولوشنی و شبکه‌های بازگشتی، یک مسئله دسته‌بندی داده‌های سری زمانی را بررسی خواهیم کرد.

از ماژول pyts در زبان پایتون می‌توانید برای دسترسی به داده‌های بازار بورس تهران استفاده کنید. با استفاده از این ماژول، سابقه روزانه شاخص کل بورس ایران و شاخص هم وزن را استخراج و به یک سری زمانی با دو متغیر تبدیل کنید. این کار را می‌توانید به

¹ Transformer

صورت دستی یا با استفاده از هر ابزار دیگری انجام دهید. هدف شما در این مسئله پیش‌بینی مثبت یا منفی بودن شاخص کل برای روز کاری آینده است. بنابراین باید ستون دیگری به عنوان برچسب که شامل دو مقدار ۱ و -۱ است به دیتاست اضافه کنید.

الف) پیش از شروع آموزش مدل، باید بر روی دیتاست عمل **Data windowing** اعمال شود. به صورت خلاصه در مورد لزوم این عمل توضیح دهید. تعداد گام‌های ورودی (N) مدل را می‌توانید به دلخواه یا با سعی و خطا تعیین کنید.

ب) یک مدل شبکه عصبی بازگشتی ساده برای پیش‌بینی وضعیت شاخص کل آموزش دهید و ارزیابی کنید. توجه کنید که مدل شما باید شاخص کل و شاخص هم‌وزن را در N روز گذشته به عنوان ورودی دریافت کرده و وضعیت شاخص کل را برای روز آینده به عنوان یک مسئله دسته‌بندی دو کلاسه (مثبت یا منفی) پیش‌بینی می‌کند.

ج) قسمت قبل را با یک شبکه کانولوشنی تکرار و مقایسه کنید.

توجه: پیش‌بینی شاخص کل تحت تاثیر عوامل بسیار زیادی است. حتی قدرتمندترین شبکه‌های موجود همچنان در پیش‌بینی بازارهای مالی بسیار با ثبات‌تر، ضعیف هستند. بنابراین کسب نتیجه ضعیف برای حل این مسئله، دور از انتظار نیست.

بخش دوم) آشنایی با شبکه خودکدگذار و کاربردهای آن

یک مسئله مهم دیگر از تحلیل سری‌های زمانی، تشخیص ناهنجاری^۲ در آن‌ها است. منظور از ناهنجاری، نقاط یا محدوده‌ای از سری زمانی است که با توجه به الگوی نقاط همسایه‌اش پیش‌بینی پذیر نیست. روش‌های مختلفی برای تشخیص ناهنجاری در سری‌های زمانی یا دیگر انواع داده‌ها ارائه شده است که یکی از رایج‌ترین آن‌ها، استفاده از شبکه‌های خودکدگذار^۳ است.

د) با توجه به [این لینک](#)، فرآیند تشخیص ناهنجاری در سری‌های زمانی را با استفاده از شبکه خودکدگذار به تفصیل توضیح دهید.

کاربرد شبکه‌های خودکدگذار تنها در تشخیص ناهنجاری خلاصه نمی‌شود. پیش از این، از این شبکه‌ها برای یادگیری ویژگی بدون نظارت^۴ استفاده شده است. علاوه بر این، اخیراً ثابت شده است که این شبکه‌ها حتی می‌توانند از طریق یادگیری خودنظارتی^۵، کارایی شبکه‌های عصبی را به شدت افزایش دهند.

ه) به عنوان یک نمونه از این کاربرد، توضیح دهید [این مقاله](#)^۶، چگونه از شبکه‌های خودکدگذار برای یادگیری خودنظارتی استفاده می‌کند. کافی است ایده اصلی این مقاله را در یک بند توضیح دهید؛ بنابراین نیاز به ذکر جزئیات مقاله نیست.

² Anomaly detection

³ Autoencoders

⁴ Unsupervised feature learning

⁵ Self-Supervised Learning

⁶ He, Kaiming, et al. "Masked autoencoders are scalable vision learners." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.