



مینی‌پروژه شماره دو

چکیده و موعد تحویل مینی‌پروژه

- برای مینی‌پروژه ملزم به ارائه گزارش متنی شامل توضیحات کامل هر قسمت هستید.
- موعد تحویل این تمرین، ساعت ۲۳:۵۹ روز جمعه مورخ ۱۴۰۲/۰۹/۲۴ است.

۱ سوال اول

مجموعه داده مربوط به این سوال را از طریق [این پیوند](#) دانلود کنید و در مراحل بعدی از آن استفاده کنید. ستون اول و دوم فایل CSV مربوط به این مجموعه داده، مربوط به ویژگی‌ها و ستون سوم آن مربوط به کلاس هر داده است.

۱. داده‌ها را با نسبت ۸۰ به ۲۰ درصد به دو قسمت آموزش و آزمون تقسیم کنید. سپس با استفاده از قاعده پرسپترون، یک نورون روی داده‌های مجموعه آموزشی، آموزش دهید (آستانه را دلخواه در نظر بگیرید).

۲. نتیجه را روی داده‌های مجموعه آزمون نشان دهید و دقت را به دست آورید. برای داده‌های تست دوخط موازی جداکننده به دست آمده از قاعده پرسپترون را نمایش دهید و داده‌های تفکیک شده دو کلاس را با رنگ مجزا در Scatter Plot مشخص کنید.

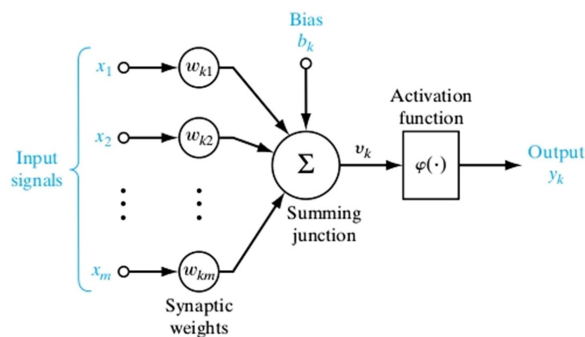
۳. قسمت‌های «۱» و «۲» را با آستانه دیگری انجام داده و نتایج را با حالت قبل مقایسه کنید. تحلیل کنید که انتخاب آستانه در پرسپترون چه تأثیری روی نتایج طبقه‌بندی دارد. ضمن پیاده‌سازی تحلیل کنید که حذف بایاس چه تأثیری بر نتایج خواهد گذاشت.

۲ سوال دوم

۱. به کمک نورون McCulloch-Pitts توسعه یافته^۱، یک ضرب‌کننده باینری بسازید که دو ورودی دوبیتی را گرفته و آن‌ها را ضرب کند. برای این کار به دو ورودی دوبیتی (در واقع چهار نورون برای همه ورودی‌ها) نیاز داریم. هم‌چنین چهار بیت خروجی (چهار نورون) مورد نیاز است. توجه شود که تمامی نورون‌های ورودی و خروجی باینری هستند (صفر و یک). ترتیب زمانی انجام عملیات در این سوال مهم نیست؛ بنابراین، نیازی به در نظر گرفتن تأخیر برای انجام عملیات نیست.

ضمن رسم جدول ورودی-خروجی، شبکه هر خروجی را به همراه توضیحات مختصری رسم کنید (نیازی به کدنویسی در این قسمت نیست). دقت داشته باشید که شبکه‌ای که برای هر خروجی رسم می‌کنید تا حد ممکن دارای کم‌ترین تعداد نورون و کم‌ترین آستانه باشد (تعداد نورون کمتر دارای اهمیت بالاتری نسبت به آستانه کوچک‌تر است). هم‌چنین توجه کنید که تمام شبکه برای یک خروجی دارای آستانه یکسان باشد.

^۱ نمایی از یک نورون McCulloch-Pitts در شکل ۱ نشان داده شده است.



شکل ۱: نمونه نورون McCulloch-Pitts.

۲. با استفاده از زبان پایتون شبکه‌های طراحی شده در قسمت «۱» را پیاده‌سازی کرده و تمامی حالات ممکن را به صورت مناسبی نشان دهید.

۳ سوال دوم

به این دفترچه‌کد مراجعه کنید و با اجرای سلول اول، ۵ داده تصویری مربوط به حروف الفبای فارسی که در شکل ۲ نشان داده شده است را دریافت کنید و سپس به سوالات زیر پاسخ دهید. دقت داشته باشید که در هر مرحله ارائه توضیحات متنی و دیداری مناسب لازم است. مثلاً می‌توانید ورودی نویزی و خروجی پیش‌بینی شده را در یک تصویر در کنار هم قرار دهید.



شکل ۲: نمونه داده‌ها.

۱. دو تابع پایتونی در سلول‌های دوم و سوم این دفترچه‌کد نوشته شده‌اند. اولین تابع تصویر را در ورودی خود دریافت و به صورت نمایش باینری درمی‌آورد و دومین تابع با افزودن نویز به داده‌ها، داده‌های جدید نویزی تولید می‌کند. در مورد نحوه عملکرد هریک از این توابع توضیح دهید. همچنین، می‌توانید این دستورات را به صورتی بهتر و کارآمدتر بازنویسی کنید.

۲. یک شبکه عصبی (همینگ یا هافیلد) طراحی کنید که با اعمال ورودی دارای میزان مشخصی نویز برای هر یک از داده‌ها، خروجی متناسب با آن داده نویزی را بیابد. میزان نویز را تا حدی که شبکه شما ناموفق عمل کند افزایش دهید و نتایج را مقایسه و تحلیل کنید.

۳. با الهام گرفتن از تابع نوشته شده برای تولید داده‌های نویزی، یک تابع بنویسید که از داده‌های ورودی، خروجی‌های دارای Missing Point تولید کند. سپس عملکرد شبکه خود را با مقدار مشخصی Missing Point آزمایش و تحلیل کنید. اگر میزان Missing Point از چه حدی بیشتر شود عملکرد شبکه طراحی شده شما دچار اختلال می‌شود؟ راه حل چیست؟ (راهنمایی: نمونه داده دارای Missing Point در شکل ۳ نشان داده شده است).



شکل ۳: نمونه داده دارای Missing Point.

۴ سوال سوم

یک مجموعه داده برای پیش‌بینی قیمت خانه‌ها را از طریق [این پیوند](#) دانلود کنید و مراحل ذکر شده در سوالات بعدی را برای فایل data.csv آن انجام دهید. لازم است که هر قسمت و مورد خواسته شده را با استفاده از دستورات پایتون انجام دهید و در جاهایی که نیاز است، نتایج را به صورت دقیق و کامل نمایش داده و تحلیل کنید.

۱. فایل CSV مربوط به این سوال را خوانده و سپس تابع Info. را از Pandas فراخوانی کنید. تعداد داده‌هایی که Nan هستند را برحسب هر ستون نمایش دهید و اگر نیاز است دستوراتی برای رفع این مشکل بنویسید.
۲. ماتریس هم‌بستگی^۱ را رسم کنید. چه ویژگی‌ای با قیمت هم‌بستگی بیش‌تری دارد؟
۳. نمودار توزیع قیمت و نمودار قیمت و ویژگی‌ای که هم‌بستگی زیادی با قیمت دارد را رسم کنید.
۴. ستون Date را به دو ستون ماه و سال تبدیل کنید و این ستون را از دیتافریم حذف کنید.
۵. داده‌ها را با نسبت ۸۰ به ۲۰ درصد به مجموعه‌های آموزش و آزمون تقسیم کنید و داده‌های آموزشی و آزمون را با استفاده از MinMaxScaler مقیاس کنید.
۶. یک مدل Multi-Layer Perceptron (MLP) ساده با ۲ لایه پنهان یا بیش‌تر بسازید. بخشی از داده‌های آموزش را برای اعتبارسنجی کنار بگذارید و با انتخاب بهینه‌ساز و تابع اتلاف مناسب، مدل را آموزش دهید. نمودارهای اتلاف و R2 Score مربوط به آموزش و اعتبارسنجی را رسم و نتیجه را تحلیل کنید.
۷. فرآیند سوال قبل را با یک بهینه‌ساز و تابع اتلاف جدید انجام داده و نتایج را مقایسه و تحلیل کنید.
۸. پنج داده را به صورت تصادفی از مجموعه ارزیابی انتخاب کرده و قیمت پیش‌بینی شده را به همراه قیمت واقعی نشان دهید. قیمت پیش‌بینی شده با قیمت واقعی چقدر تفاوت دارد؟ آیا این عملکرد مناسب است؟ برای بهبود آن چه پیشنهادی دارید؟

۵ سوال پنجم

۱. مجموعه داده Iris را فراخوانی کنید و روش‌های تحلیل داده‌ای که آموخته‌اید را روی آن به کار ببندید. داده‌ها را با نسبتی دلخواه و مناسب به مجموعه‌های آموزش و ارزیابی تقسیم کنید.
۲. با استفاده از روش‌های آماده پایتون، سه مدل بر مبنای رگرسیون لجستیک، MLP و شبکه‌های عصبی پایه شعاعی (RBF) را تعریف کرده و روی داده‌ها آموزش دهید. نتایج روی داده‌های ارزیابی را حداقل با چهار شاخص و ماتریس درهم‌ریختگی نشان داده و تحلیل کنید. در انتخاب فرایارامترها آزاد هستید؛ اما لازم است که نتایج را به صورت کامل مقایسه و تحلیل کنید. به دانشجویانی که این سوال را بدون استفاده از کتابخانه‌ها و مدل‌های آماده پایتونی انجام دهند، تا ۲۰ درصد نمره امتیازی تعلق خواهد گرفت.

منابع

[1] <https://github.com/MJAHMADEE/MachineLearning2023>

¹Correlation Matrix