



۱. (۲۵٪) [مطالعه و بررسی] توابع فعال سازی (Activation functions) زیر را که بر مبنای تاریخ توسعه مرتب شده‌اند، در نظر بگیرید:

- Sigmoid
- Tanh
- Softmax
- Rectified Linear Unit (ReLU)
- Leaky ReLU
- PReLU
- Exponential Linear Unit (ELU)
- Maxout
- Gelu
- Swish
- Mish

۱-۱ نمودار هر یک از آنها را به ازای متغیر x در بازه $x \in [-5, 5]$ رسم کنید.

۲-۱ تابع گرادینان هر کدام از توابع را نسبت به متغیر x در بازه $x \in [-5, 5]$ بدست آورید و آنها را رسم کنید.

۳-۱ کدام یک از توابع فعال سازی بالا مناسب برای مسائل دوکلاسه (Binary Classification) و کدام یک مناسب برای مسائل چندکلاسه (Multi Class Classification) مناسب هستند؟

۲. (۷۵٪) [پیاده سازی: پرسپترون و آدالین] مسئله بازشناسی نویسه را برای الگوهای بیان شده در مثال ۲-۱۵ در فصل دوم کتاب، با استفاده از ساختار شبکه نشان داده شده در شکل ۲-۲۱ پیاده سازی کنید. بدین منظور الگوهای آموزش شکل ۲-۲۰ و الگوهای آزمون (تست) شکل ۲-۲۲ که به صورت فایل های متنی به همراه تمرین ارائه شده است، به کار بگیرید. برای موارد زیر نتیجه را گزارش کنید. در گزارش ها درصد خطای بازشناسی را به صورت زیر محاسبه و گزارش کنید.



$$Error Rate = \frac{N_{err}}{N} \times 100 = \frac{\text{تعداد الگوهایی که اشتباهی بازشناسی شده اند}}{\text{تعداد کل الگوها}} \times 100$$

نکته: برای بهتر دیدن الگوهای ارائه شده در فایل‌های متنی، می‌توانید از فونت Courier New یا CourierPS در NotePad استفاده کنید.

۱-۲ (۳۰٪) [پرسپترون] آموزش شبکه را با استفاده از قانون پرسپترون و با در نظر گرفتن موارد خواسته شده در بخش‌های زیر انجام دهید. توجه شود در هر بخش جدید، بهترین پارامترهای پیدا شده در بخش‌های پیشین را به عنوان مقادیر پیش فرض در نظر بگیرید:

الف) [مقادیر اولیه] در این بخش قصد داریم درصد خطای بازشناسی را برای داده‌های آزمون، به ازای اجزای مختلف از مقادیر مختلف اولیه وزن‌ها و بایاس‌های شبکه پیدا کنیم. در این راستا این مقادیر را با روش‌های زیر تولید کرده و سپس درصد خطای بازشناسی و زمان همگرایی را برای داده‌های آموزش گزارش کنید. با استفاده از نتایج، در مورد اثر مقادیر اولیه بر کارایی و سرعت شبکه بحث کنید. روش‌های تولید وزن اولیه:

۱- همه صفر

۲- مقادیر تصادفی با توزیع یکنواخت

۳- مقادیر تصادفی با توزیع نرمال

۴- روش Xavier(Glorot)

۵- روش Kaiming(HE)

ب) [مقدار آستانه] برنامه را با بهترین روش بخش الف، برای چند مقدار مختلف آستانه θ تکرار کنید و نتایج بازشناسی را برای داده‌های آزمون ارائه کنید. آیا مقدار بیشتر θ تأثیری بر درصد خطای شبکه دارد؟

ج) [نرخ یادگیری] شبکه را با بهترین روش بخش الف و ب، برای پنج مقدار مختلف از نرخ یادگیری، ۰.۵، ۰.۱، ۰.۰۵ و ۰.۰۱ آموزش دهید و نتیجه بازشناسی بر روی داده‌های آزمون را به صورت نمودار رسم کنید. زمان همگرایی شبکه را در هر حالت گزارش کنید. با توجه به نتایج



حاصل، در مورد تاثیر این پارامتر بر عملکرد شبکه بحث کنید.

۲-۲ (۲۰٪) [آدالاین] شبکه را با استفاده از قانون آدالاین (دلتا) و با بهترین پارامترهای پیدا شده در قسمت‌های قبل سوال (نرخ یادگیری، وزن‌های اولیه و بایاس) آموزش دهید و نتیجه بازشناسی را برای داده‌های آزمون و داده‌های آموزش گزارش کنید.

۳-۲ (۵٪) [مقایسه] بهترین نتایج بدست آمده را روی هر دو مجموعه آموزش و آزمون در ۳ پیاده‌سازی مربوط به پرسپترون و پیاده‌سازی مربوط به آدالاین، در یک جدول بیاورید. نتایج را مقایسه کرده و علل تفاوت نتایج را شرح دهید.

۴-۲ (۱۰٪) [داده نویزی] برای حالتی که برای الگوی تست، بیش از یکی از دسته‌ها انتخاب می‌شوند، چه راهکاری را پیشنهاد می‌کنید. روش خود را پیاده‌سازی کرده و نتیجه را برای شبکه پرسپترون گزارش کنید.

۵-۲ (۱۰٪) [استخراج ویژگی] در بخش‌های قبل از این تمرین، مقدار کل پیکسل‌های هر نویسه به عنوان ورودی شبکه استفاده شد. در این بخش، از روش تصویر کردن (projection) برای استخراج ویژگی استفاده می‌شود و مقدار ویژگی‌ها (به جای مقادیر پیکسل‌ها) به عنوان ورودی به شبکه داده می‌شود. در این روش، به ازای هر ردیف (و هر ستون) از هر نویسه، مجموع پیکسل‌های روشن (با مقدار یک) آن ردیف (یا ستون) شمارش شده و مقدار حاصل جمع به عنوان ویژگی در نظر گرفته می‌شود. با توجه به ابعاد نویسه‌ها که 9×7 هستند، تعداد ویژگی‌های هر نویسه $9 + 7 = 16$ خواهد بود.

شبکه آدالاین را برای آموزش با ویژگی‌های حاصل تغییر دهید و نتایج حاصل را (نرخ خطا روی مجموعه آزمون) در یک جدول گزارش کنید. آیا کارایی شبکه‌ها در این حالت، به نسبت حالتی که مقدار خود پیکسل‌ها استفاده شود، بهبود می‌یابد یا خیر؟ در هر صورت، تحلیل خود را از این نتایج بیان کنید.