



۱. (۸٪) [پروژه درس] گزارشی از پیشرفت پروژه درس را بیان کنید.

۲. (۲۲٪) [پژوهش] تاکنون پروژه‌های مختلفی در حوزه یادگیری عمیق برای پردازش گفتار پیاده‌سازی و ارائه شده است (مثلاً Tacatron برای تبدیل متن به گفتار و Wav2Vec برای تبدیل گفتار به متن). گزارشی از مطالعه این پروژه‌ها برای سه حوزه بازشناسی گفتار، تبدیل متن به گفتار و بهسازی گفتار ارائه دهید که در آن برای هر حوزه حداقل دو پروژه را پوشش دهد. در گزارش خود، علاوه بر بیان مختصر مشخصات آن پروژه، مشخص کنید هر کدام از آنها کدامیک از الگوریتم‌های یادگیری ماشین را دارا استفاده کرده‌اند و برای هر حوزه، مقایسه مختصری از پروژه‌ها با همدیگر ارائه دهید.

۳. (۷۰٪) [پیاده‌سازی: تشخیص اعداد انگلیسی با شبکه عصبی MLP] از شبکه عصبی پرسپترون چندلایه (MLP) برای تشخیص اعداد انگلیسی صفر تا نه استفاده کنید. برای این کار از دادگان صوتی تمرین قبل استفاده کنید. پیاده‌سازی شبکه باید توسط خودتان انجام شده باشد و استفاده از ابزارهای آماده مجاز نیست. تعداد نرون‌های ورودی شبکه را برابر با تعداد کل ویژگی‌های فایل، و تعداد نرون‌های خروجی را به تعداد دسته‌ها (۱۰ دسته) در نظر بگیرید. تعداد نرون‌های میانی را تقریباً میانگین تعداد نرون‌های ورودی و خروجی در نظر بگیرید. مقدار نرخ یادگیری را 0.001 قرار دهید و تابع فعال‌سازی لایه میانی و خروجی را سیگموئید دوقطبی قرار دهید. از کل مجموعه داده آموزش برای یادگیری شبکه و از داده آزمون برای ارزیابی کارایی شبکه استفاده کنید.

الف) از روش MFCC با طول فریم ۲۰ میلی‌ثانیه، ۲۴ فیلتر مل و تعداد ۱۲ ویژگی به همراه مشتق‌های اول و دوم استفاده کنید و مقدار Accuracy هر عدد و میانگین دقت کلی را گزارش کنید. همچنین نمودار خطای MSE برحسب تکرار را رسمی کنید.

برای اینکه کار پیاده‌سازی آسان شود، همه داده‌ها را با افزودن صفر به اول و آخر فایل‌ها با بزرگترین نمونه (در کل داده‌های آموزش و آزمون) هم طول کنید. به نظر شما راه دیگری برای حل این سوال و ثابت کردن تعداد نرون‌های ورودی وجود دارد؟ چه راهی؟

ب) تعداد لایه‌های مخفی را به دو لایه افزایش دهید و نتایج بخش الف را برای دو حالت مختلف،



بدست آورید. در یک حالت تعداد نرون‌های دو لایه مخفی به ترتیب، برابر با نصف و یک سوم تعداد نرون ورودی باشد و در حالت دیگر، تعداد نرون‌های این دو لایه برابر با دو سوم و نصف باشد. برای همه نرون‌های این دولایه از تابع فعال‌سازی سیگموئید استفاده کنید. در مورد مقایسه نتایج بخش الف و ب چه نتیجه‌ای می‌توان گرفت؟

۴. (۴۰٪ نمره اضافی) [پیاده‌سازی: تشخیص اعداد انگلیسی با شبکه عصبی CNN] در این تمرین می‌خواهیم از شبکه عصبی پیچشی (CNN) برای تشخیص اعداد انگلیسی (مشابه سوال قبل) استفاده کنیم. برای این کار، می‌توانید از کتابخانه‌ها و بسترهای آماده استفاده کنید. در ابتدا لازم است که فایل‌های گفتاری اعداد به تصویر تبدیل شوند که برای این منظور از طیف‌نگار (Spectrogram) استفاده کنید (لزومی به هم طول کردن فایل‌ها نیست). سپس تصاویر دادگان آموزش را برای Fine Tune کردن شبکه AlexNet استفاده کنید. برای این کار، لایه آخر آن را حذف کرده و ۱۰ نرون، معادل تعداد اعداد، با وزن‌های اولیه تصادفی قرار دهید و برای سایر لایه‌های از همان وزن‌های آموزش داده شده AlexNet به عنوان پیش آموزش استفاده کنید.

دقت این روش با دقت بدست آمده در سوال قبلی مقایسه کنید و تحلیل خود را درباره آن بیان کنید.