

مینی پروژه درس یادگیری ماشین

هدف از انجام این مینی-پروژه، بررسی روش‌های یادگیری نظارت شده بر روی دیتاست MNIST است. دیتاست MNIST یک دیتاست بسیار مشهور و قدیمی است. در این دیتاست ۶۰۰۰۰ تصویر آموزشی و ۱۰۰۰۰ تصویر آزمون قرار دارد. هر یک از این تصاویر، یک تصویر 28×28 پیکسل بریده شده و تمیز شده است که یک عدد دست نویس در داخل آن نوشته شده است. برای انجام این تمرین، تمام تنظیم پارامترها و مقایسه مدل‌ها را به کمک بخش validation از دیتاست آموزشی انجام دهید و در آخر، دقت بهترین مدل خود را بر روی دادگان آزمون گزارش کنید. (۶۰۰۰۰ داده آموزشی را به دو بخش دادگان آموزشی با ۴۸۰۰ داده و validation با ۱۲۰۰ داده تقسیم کنید دادگان را پیش از تقسیم کردن حتماً بر بزنید. سپس آموزش و ارزیابی مدل را بر روی این دادگان انجام دهید و مدل نهایی خود را بر روی دادگان آزمون اصلی تست کنید و نتیجه را گزارش کنید.)

برای یافتن بهترین مدل، مدل‌های Logistic، Regression، SVM، KNN، Neural Network و را با یکدیگر مقایسه کنید. در هر یک از این مدل‌ها تلاش کنید بهترین مجموعه هایپر پارامترها را بیابید که دقت بیشتری را تولید کند. همچنین استفاده از هر تکنیک دلخواه برای بالا بردن دقت، با ذکر دلیل استفاده مجاز است و نمره اضافی در بر خواهد داشت.

برای انجام این پروژه توصیه می‌شود به صورت دلخواه خود عمل کنید اما می‌توانید از مراحل زیر استفاده کنید:

۱. بررسی ورودی‌ها و خروجی‌ها. تلاش کنید تا تصویر را به صورت تصاویر 28×28 پیکسلی نمایش دهید و با لیبیل مقایسه کنید.

۲. نرمالیزیشن

۳. بررسی روش‌های Logistic، Regression، KNN، SVM، Decision Tree، Random Forest، Neural Networks برای هر کدام از این روش‌ها، تلاش کنید تا بهترین دقت ممکن را با تنظیم هایپر پارامترها به دست بیاورید. سپس دقت نهایی خود را بر روی دیتاست validation برای بهترین حالت یافته شده در هر مدل مقایسه کنید تا بهترین مدل را بیابید.

۴. توصیه می‌شود در نهایت و پس از تعیین هایپر پارامترها، برای هر مدل وجود overfitting و underfitting را نیز بررسی کنید.

۵. مدل نهایی را بر روی دیتاست آزمون که از اول کنار گذاشته بودید ارزیابی کنید. مقادیر accuracy, recall, precision را محاسبه کنید و گزارش کنید. همچنین confusion matrix را برای خروجی مدل نهایی محاسبه کنید و گزارش کنید.

- در نمره دهی به تمارین، معیار بیشتری دقت نخواهد بود بلکه معیار بررسی مناسب مدل‌ها خواهد بود. لذا نگران یافتن دقتی کمتر از دقت بقیه نباشید.

- توصیه می‌شود تا برای بهتر متوجه شدن تکنیک‌های لرنینگ، ایرادات، قدرت و ضعف‌های الگوریتم‌ها و همچنین تسلط بیشتر، خروجی‌های مدل‌های خود را دقیقاً بررسی کنید، تصاویری که منجر به خطا شده‌اند ببینید، از تکنیک‌های visualization کمک بگیرید و البته استفاده از این روش‌ها اجباری نیست.

- می‌توانید روش‌های bagging و boosting را نیز بر روی این دیتاست بررسی کنید اما به علت زیاد بودن هایپرپارامترها، بررسی آنها اجباری نیست.

- پس از انجام روندهای مطرح شده، نمودارهای نهایی و دقت‌ها و ... را به ترتیب به همراه یک عنوان در یک فایل قرار دهید و اگر توضیحاتی دارید بنویسید. نیازی به صرف زمان زیاد بر روی گزارش نیست.

تاریخ تحویل پروژه: جمعه ۱۴ شهریور تا ساعت ۱۲ شب. گزارش خود را به صورت فایل PDF در ایمیل زمیمه کنید، کدهای خود را در github آپلود کرده و آدرس github خود را در ایمیل قرار دهید و سپس برای بنده ارسال کنید. همچنین عنوان نامه را به صورت Name-ML-MiniProject قرار دهید.

mohamadqodosi۱۹۹۶@gmail.com