

بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران

پائیز ۱۳۹۷

تمرین سری ۷

تحويل: شنبه ۱۷ آذر

۱. برنامه‌ای بنویسید که با استفاده از یک طبقه‌بند ساده اعداد دست نویس فارسی را که داده‌های آن به پیوست ارسال شده و از طریق لینک زیر قابل دانلود است، طبقه‌بندی کند.

<https://github.com/amir-saniyan/HodaDatasetReader>

ویژگی‌ها:

پیکسل‌های تصویر برای یادگیری چندان مناسب نیستند. بنابراین قبل از ورود به مرحله آموزش لازم است ویژگی‌هایی از هر تصویر استخراج شوند تا طبقه‌بند بتواند بر اساس آنها طبقه‌بندی کند. در مرحله آموزش، ویژگی‌های تصاویر و برچسب‌های مربوطه به طبقه‌بند وارد می‌شوند تا مدل آموزش ببیند و در مرحله آزمون ویژگی‌های هر تصویر توسط مدل استفاده می‌شود تا برچسب آن را پیش‌بینی کند. برای این مرحله از ویژگی‌های هندسی و البته ویژگی‌های دیگر استفاده کنید. برای استخراج ویژگی می‌توانید از کدهای آماده استفاده کنید.

فاز آموزش:

در فاز آموزش، بخشی از داده‌ها برای یادگیری پارامترهای مدل استفاده می‌شود. داده‌های آموزش شامل تصویر و برچسب است که برچسب نوع کلاس تصویر (اعداد ۰ تا ۹) را مشخص می‌کند. کد شما باید شامل این فرآیند باشد. برای طبقه‌بند می‌توانید از کدهای آماده (مانند SVM یا kNN) استفاده کنید.

فاز آزمون:

پس از اتمام فاز آموزش، می‌توان با استفاده از طبقه‌بند آموزش دیده، برچسب تصویر جدید را پیش‌بینی کرد. بخشی از داده‌های دیتاست داده شده مربوط به آزمون و بخشی دیگر مربوط به فاز آموزش هستند. کد شما باید شامل مرحله آزمون نیز باشد.

ارزیابی طبقه‌بند:

برای ارزیابی طبقه‌بند از ماتریس confusion و دقت میانگین استفاده کنید. این شاخص‌ها میزان کارایی طبقه‌بند شما را نشان می‌دهد که تا چه میزان داده‌های آزمون را درست پیش‌بینی کرده است.

گزارش کاملی از مراحل انجام شده و نتایج بدست آمده را در سایت درس همراه با کدهای مربوطه بارگذاری کنید (در قالب یک فایل فشرده). مقایسه نتایج حاصل از ویژگی‌های مختلف و طبقه‌بندهای مختلف دارای امتیاز است.