

تمرین سری چهارم درس یادگیری ماشین

در این تمرین، می‌خواهیم روش‌های خوشه‌بندی را بر روی دادگان MNIST بررسی کنیم. مجموعه دادگان MNIST در مینی پروژه اول در اختیار شما قرار گرفته است. برای انجام این تمرین، دادگان MNIST را در یک ماتریس نامپای ذخیره کنید، سپس اقدام به خوشه‌بندی آن به کمک الگوریتم DB-Scan کنید و تعداد خوشه‌ها، این که در خوشه‌ها چه دادگانی قرار گرفته‌اند، این که آیا این خوشه‌ها مناسب کلاس بندی هستند یا نه و این که داده‌هایی که به عنوان noise تشخیص داده شده‌اند چه تعداد هستند و چه داده‌هایی نویز تشخیص داده شده‌اند را بررسی کنید. برای تنظیم پارامترهای DB-Scan پارامترهای مختلف را آزمایش کنید و پارامتری که به نظر شما مناسب‌تر است (تعداد نویزها زیاد نباشند و تعداد خوشه‌ها هم بین ۱۰ تا ۲۰ باشد) را انتخاب کنید. در بخش بعدی تمرین، به کمک روش k-means دادگان را به تعدادی خوشه تقسیم کنید. برای تعداد خوشه‌ها، یک بار ۱۰ خوشه در نظر بگیرید و دقت گروه‌بندی را حساب کنید و در دفعه دوم، تعداد خوشه‌ها را به کمک روش توضیح داده شده (elbow) محاسبه کنید. برای بررسی دقت گروه‌بندی، در هر خوشه، مشابه برگ‌های درخت تصمیم عمل کنید. یعنی در هر خوشه، کلاسی که بیشترین تکرار را دارد حساب کنید و برای تمام دادگان آن خوشه، همین کلاس را پیش‌بینی کنید. با این روش می‌توانید دقت گروه بندی را در خوشه‌ها حساب کنید. در نهایت، در کد خود که در گیت‌هاب خود آپلود می‌کنید، تمام نمودارهای لازم را رسم کنید و پاسخ خود به سوالات مطرح شده را نیز در ایمیل قرار داده و برای بنده ارسال کنید.

mohamadqodosi۱۹۹۶@gmail.com