



درس : مباحث ویژه در داده‌کاوی

دانشجو : امیرمحمد خرازی

شماره دانشجویی : ۴۰۱۵۲۵۲۱۰۰۲

استاد درس : دکتر منصور رزقی آهق

دانشکده علوم ریاضی ، گروه علوم کامپیوتر، گرایش داده‌کاوی

تمرین دوم

گیت‌هاب این تمرین (لینک)

گیت‌هاب درس (لینک)

اطلاعات اضافی

برای اطلاعات بیشتر لطفاً به گیت‌هاب مراجعه نمایید:

گیت‌هاب

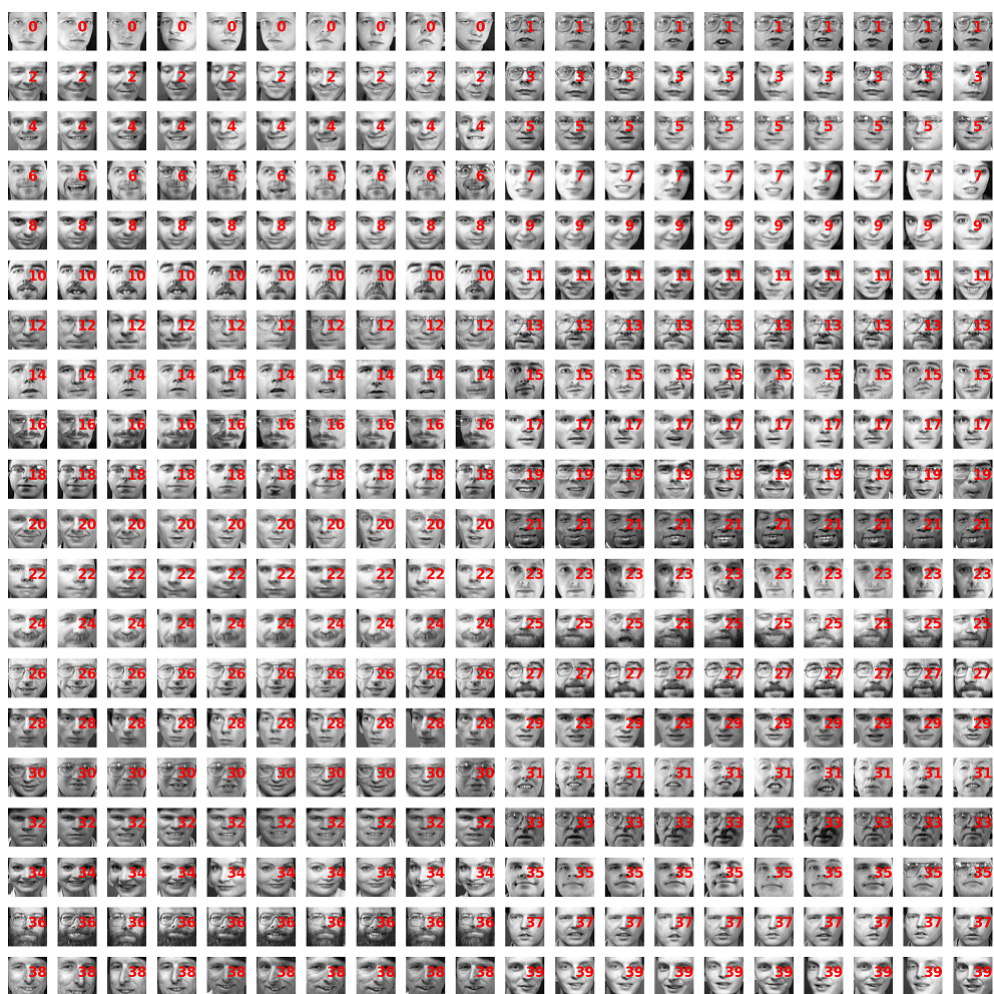
روش و کد به اندازه توان گزارش نویسی شده است و نتایج هر کدام از الگوریتم‌ها در آن مشخص است :
در مورد الگوریتم Spectral Clustering ، این الگوریتم را از پایه (یعنی بدون کمک sklearn) و با توجه به الگوریتم کتاب زکی، در صفحه ۴۵۹ نوشته ام .

گراف مورد نیز به دو روش RBF و KNN نیز به صورت دستی و مرحله مرحله حساب شده است.
تنها بخشی از الگوریتم Spectral Clustering که از sklearn استفاده می‌شود بخش آخر آن برای مشخص کردن خوشه‌های نهایی است.

در این میان تنها بخشی از سوال تمرین که باقی مانده است، درست کردن ماتریس شباهت با استفاده از OMP است.
بخش‌های دیگر آن یعنی بدست آوردن ماتریس لاپلاسی با استفاده از این ماتریس شباهت و ادامه آن تا بدست آوردن نتایج نهایی آورده شده است ولی از آنجایی که ماتریس شباهت بدست آمده، با omp درست نشده است، این بخش از سوال پاسخ داده نشده است.

می‌توانید برای بررسی بیشتر، گزارش کد را مطالعه فرمائید.
اطلاعات کامل تر شامل نحوه انجام الگوریتم‌ها و دیگر همه در کد آورده شده است.
نتایج نیز در آنجا قابل مشاهده است.
بدلیل حجم زیاد فایل output های آن حذف شده است تا در گیت‌هاب بهتر قرار بگیرد.
لطفاً در صورت نیاز آن‌ها را اجرا کنید.

برای هر مدل (kmeans و غیره) و همچنین در Spectral Clustering با هر نوع گراف (rbf و غیره) تصاویری مانند شکل زیر ایجاد میشوند.



هر سلول را در کد اصلی اجرا کنید
 بعضی از خروجی ها برای کمتر کردن حجم فایل پاک شده اند