

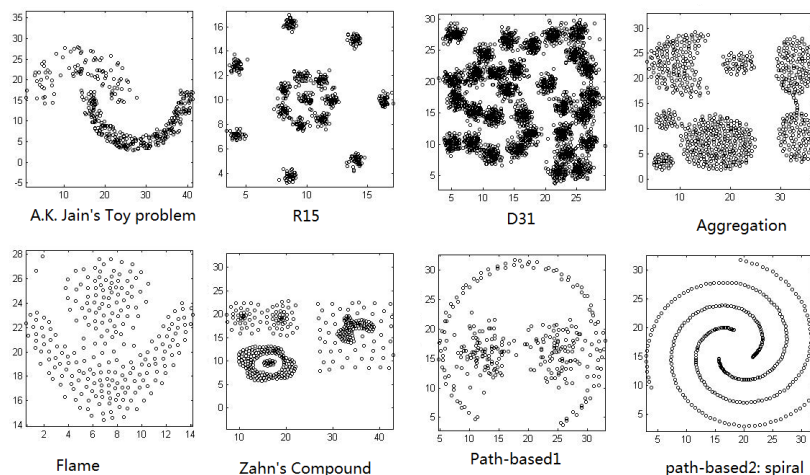
تمرین سری سوم

هدف از این تکلیف، مقایسه تجربی دو روش خوشه بندی k میانگین و خوشه بندی طیفی می باشد. برای این منظور:

۱. ابتدا مجموعه داده های قسمت shape sets را از لینک زیر دانلود کنید.

<http://cs.joensuu.fi/sipu/datasets/>

این مجموعه داده، حاوی هشت فایل متنی می باشد، که هر کدام حاوی یک سری نقاط برچسپ دار در فضای دو بعدی می باشد. هر خط از فایل شامل سه عدد می باشد. دو عدد اول مختصات نقطه و عدد سوم برچسپ نقطه را نشان می دهد. دقت کنید نقاط بدون برچسپ به الگوریتم خوشه بندی داده می شود. برچسپ ها تنها برای سنجش کیفیت خوشه بندی مورد استفاده قرار می گیرید. شکل داده ها (بدون برچسپ) در تصویر زیر نشان داده شده است.



۲. الگوریتم خوشه بندی k میانگین را پیاده سازی کرده و روی داده های مختلف قسمت بالا اجرا کنید. مراکز اولیه خوشه ها را بصورت تصادفی از جمع نقاط انتخاب کنید. مقدار k را برای هر فایل داده متناسب با تعداد برچسپ های متمایز در نظر بگیرید. برای هر اجرا زمان اجرای الگوریتم را ثبت کنید.

۳. الگوریتم خوشه بندی طیفی spectral را پیاده سازی کنید. با فرض اینکه k تعداد برچسپ ها (تعداد خوشه ها) باشد، ابتدا گراف بدون وزن k nearest neighbour را مربوطه محاسبه کرده و سپس ماتریس لاپلاسی L را تشکیل دهید. شما بایستی k بردار ویژه نخست L را بدست آورید و مطابق آنچه در کلاس توضیح داده شد، سطرهای گراف U حاصل را با استفاده از الگوریتم k میانگین خوشه بندی کنید. توجه کنید برای محاسبه بردارهای ویژه می توانید از ماژول های موجود کمک بگیرید. زمان اجرا را برای هر فایل داده ثبت کنید.

۴. نتایج بدست آمده توسط دو الگوریتم k میانگین و خوشه بندی طیفی را در جدولی مقایسه کنید. برای مقایسه از معیارهای purity (درجه خلوص خوشه ها) و F measure و زمان اجرا کمک بگیرید.