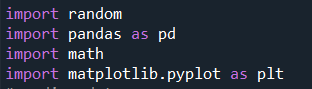
داده کاوی

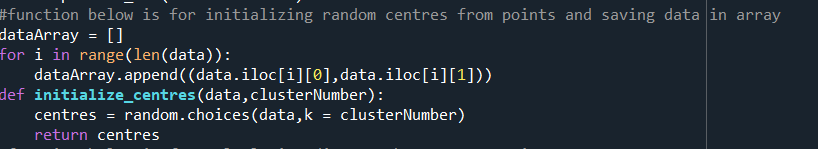
تمرین سری چهارم

سوال اول

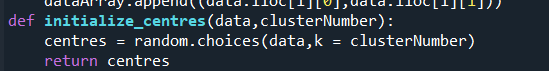
در ابتدا کتابخانه های مورد نیاز را import میکنیم



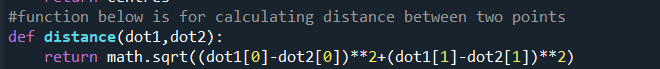
حال داده را میخوانیم و به فرم لیستی ازتاپل ها در می اوریم تا به این صورت که عضو اول هر تاپل x داده و عضو دوم y داده میباشد



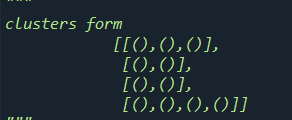
تابع زیر نیز برای مقدار دهی اولیه به مرکز ها به صورت رندوم میباشد



تابع زیر برای محاسبه فاصله دو نقطه میباشد

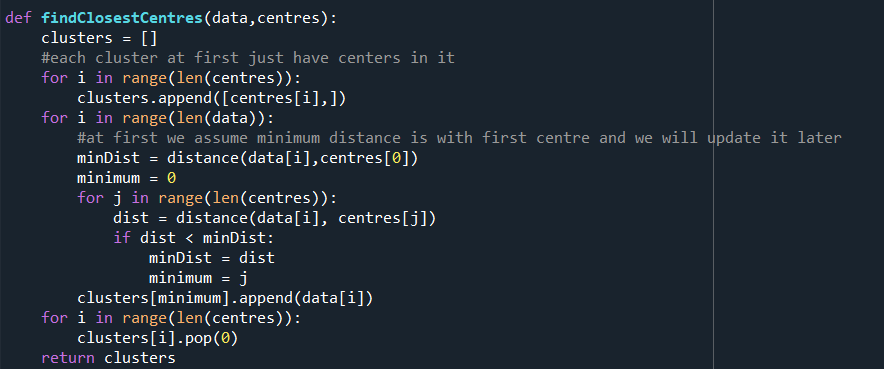


تابع findClosestCentres با گرفتن داده و مراکز خوشه ها را با اعضاش برمیگرداند که مقدار برگردانده شده به شکل زیر میباشد

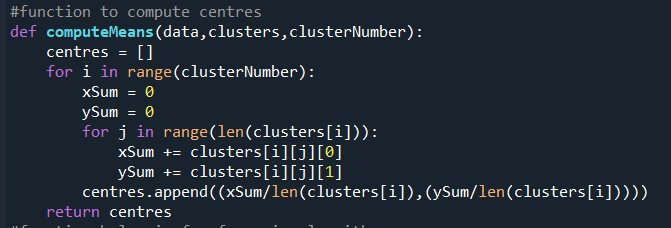


هر کلاستر یک لیست میباشد که شامل تعدادی تاپل میباشد که نشان گر نقطه متعلق به آن خوشه میباشد

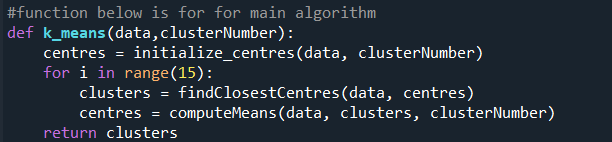
چون مقدار اولیه این کلاستر ها شامل مراکز میباشد و در محاسبه دوباره مراکز را در خوشه قرار میدهیم عضو اول هر خوشه که مرکز آن خوشه است را حذف میکنیم



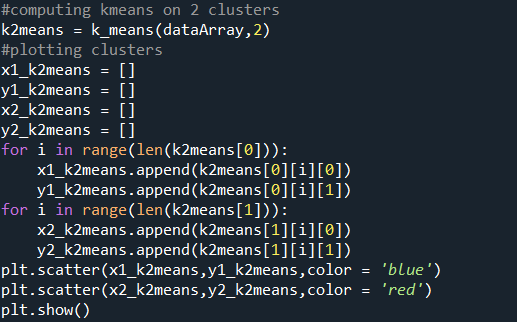
برای محاسبه مرکز با استفاده از تابع computeMeans با میانگین گیری از نقاط هر خوشه مرکز را محاسبه میکنیم

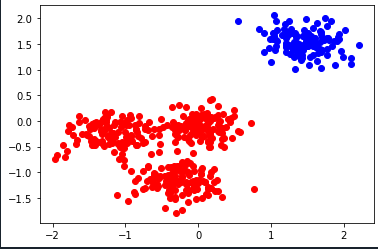


و در نهایت الگوریتم kmeans که به ازای k های مختلف جواب میدهد و حلقه ای که 15 بار اجرا میشود

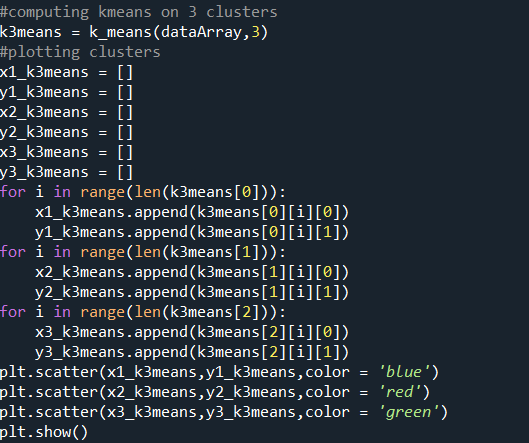


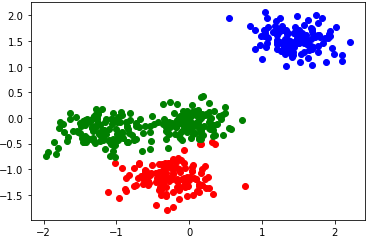
محاسبه و نمایش الگوریتم با دو خوشه



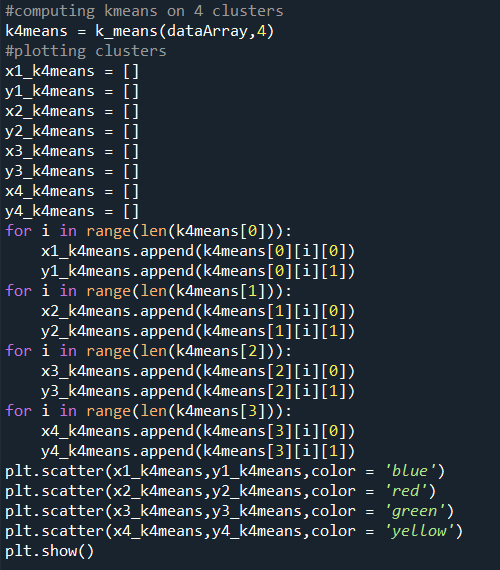


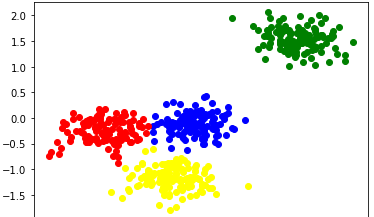
محاسبه و نمایش الگوریتم با سه خوشه





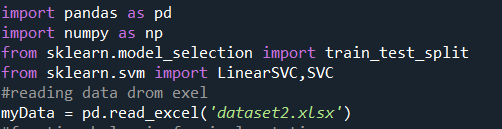
محاسبه و نمایش الگوریتم با چهار خوشه



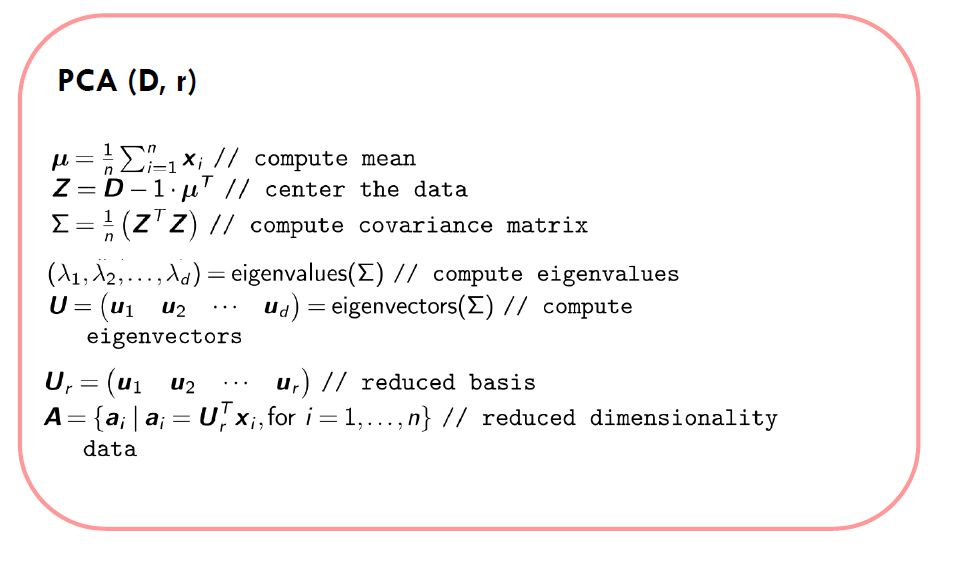
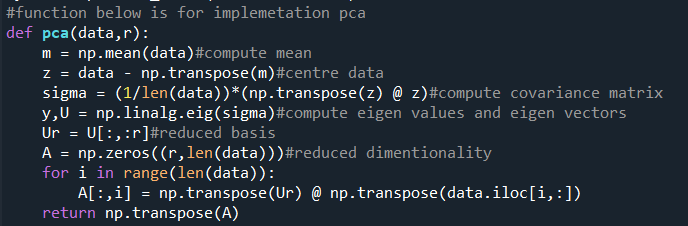


سوال دوم

در ابتدا کتاب خانه های مورد نیاز را import میکنیم و داده را میخوانیم

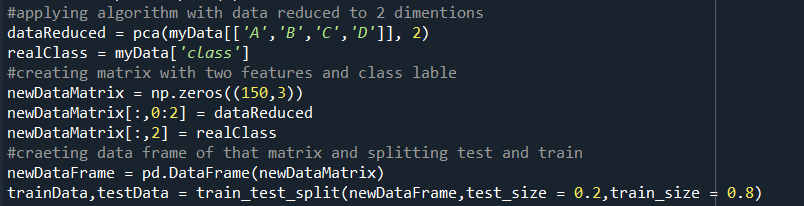


تابع زیر پیاده سازی الگوریتم pca میباشد که مطابق اسلاید ع=های درس گام به گام آن اجرا شده است

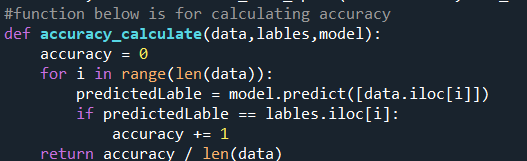


در خط اول تابع ماتریس میانگین بدست میاوریم و در خط دوم z را بدست میاوریم ماتریس کواریانس نیز بر اساس z بدست می آید و بردار و مقادیر ویژه از ماترس کواریانس با استفاده از تابع np.linalg.eig بدست می‌آید و در نهایت ماتریس A که داده های کاهش بعد یافته است را برمیگردانیم

حال داده های کاهش بعد یافته را با لیست کلاس واقعی به صورت جداگانه داریم که با ادغام آن دو ماتریس داده کاهش یافته و از روی ماتریس دیتا فریم داده های کاهش یافته را میسازیم و بعد از آن به دو بخش آموزش و تست تقسیم میکنیم



تابع زیر نیز محاسبه دقت بر اساس داده ها و مدل ماست



در نهایت یک svm خطی و یک svm غیرخطی میسازیم و دقت را برای هر دو مدل محاسبه میکنیم

