

سوال اول:

رویکرد مورد استفاده برای اینکار رویکرد onVsAll است. به این صورت که اگر یک مساله m کلاسه داشته باشیم، یک کلاس آنرا به عنوان کلاس مثبت و بقیه کلاسها را به عنوان کلاس منفی در نظر میگیریم و سپس یک دسته‌بند SVM برای آن آموزش میدهیم و سپس همین کار را برای $m-1$ کلاس دیگر تکرار میکنم. در این روش باید به اندازه تعداد کلاسها (m) دسته‌بند SVM آموزش دهیم. در مرحله تست نیز برای یک داده جدید هرکدام از دسته‌بندها جواب True (1) تولید کرد. داده مربوط به کلاس متناظر با این دسته‌بند است.

سوال دوم:

برای $k=1$ این شرط برقرار است و اگر $k>1$ باشد.

داریم $X = x_1 x_2 \dots x_k$ که X یک frequent itemset با اندازه k است. حال مقدار زیر را تعریف میکنیم:

$$X(1) = x_1 x_2 \dots x_{k-2} x_{k-1}$$

$$X(2) = x_1 x_2 \dots x_{k-2} x_k$$

به روشنی واضح است که $X = X(1) \cup X(2)$ است. همچنین جفت $X(1)$ و $X(2)$ دارای طول $k-1$ هستند و همچنین frequent هستند.

بدین ترتیب $X(1), X(2) \in F_{k-1}$ ، بنابراین X در مرحله تولید frequent itemset ها تولید میشود و همچنین در مرحله هرس حذف نمیشود. بنابراین با شروع از یک مجموعه frequent، تمام زیر مجموعه‌های آن نیز frequent هستند.

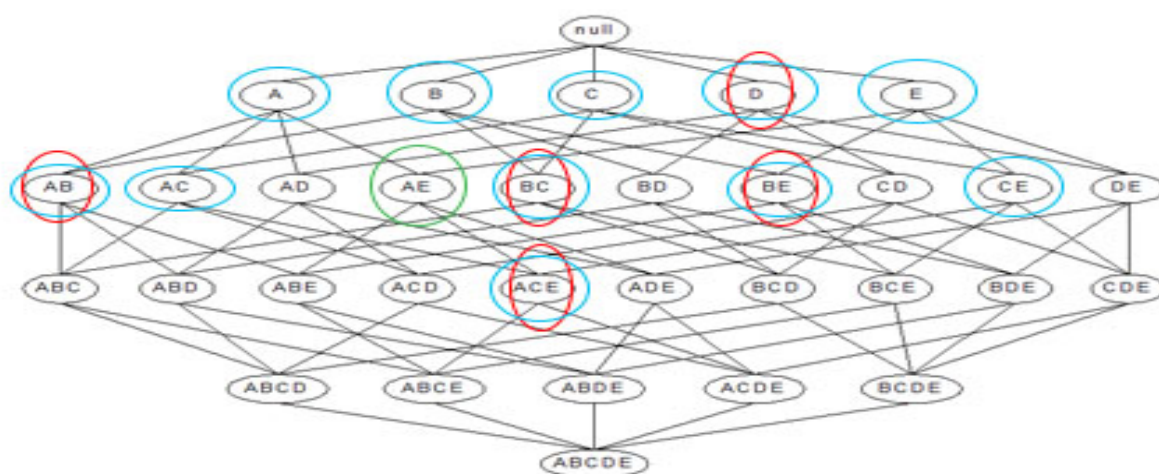
سوال سوم:

قسمت الف)

در این تمرین برای سادگی هرکدام از اقلام جدول به صورت زیر کدگذاری میشوند.

کد	نام آیتم
A	رمان
B	تاریخی
C	مذهبی
D	شعر
E	روانشناسی

نمودار این قسمت به صورت زیر است:



همچنین maximal frequent itemset ها با رنگ قرمز، close frequent itemset ها با رنگ آبی، frequent but neither maximal nor close، با رنگ سبز و بقیه که با هیچ رنگی مشخص نشده‌اند infrequent هستند.

قسمت ب)

تمام frequent itemset هایی که به رمان ختم میشوند به صورت زیر هستند.

[رمان، مذهبی]

[رمان، روانشناسی]

[رمان، مذهبی، روانشناسی]

قسمت ج)

تمام قواعد انجمنی استخراج شده از قسمت ب به صورت زیر هستند:

(روانشناسی، مذهبی)=<(رمان)

(روانشناسی، رمان)=<(مذهبی)

(مذهبی، رمان)=<(روانشناسی)

(رمان)=<(روانشناسی)

(روانشناسی)=<(رمان)

(رمان)=<(مذهبی)

(مذهبی)=<(رمان)

سوال چهارم:

در مرحله اول بین هر جفت فاصله اقلیدسی را حساب میکنیم که به صورت زیر است.

دانشجو 4	دانشجو 3	دانشجو 2	دانشجو 1	
3.60	1.73	3.31	0	دانشجو 1
4.24	2.82	0	3.31	دانشجو 2
2.44	0	2.82	1.73	دانشجو 3
0	2.44	4.24	3.60	دانشجو 4

با توجه به اطلاعات جدول بالا فاصله دانشجوی یک و سه از هم کمترین فاصله است. پس این دو دانشجو با هم ادغام میشوند و در جدولی به صورت زیر قرار میگیرند.

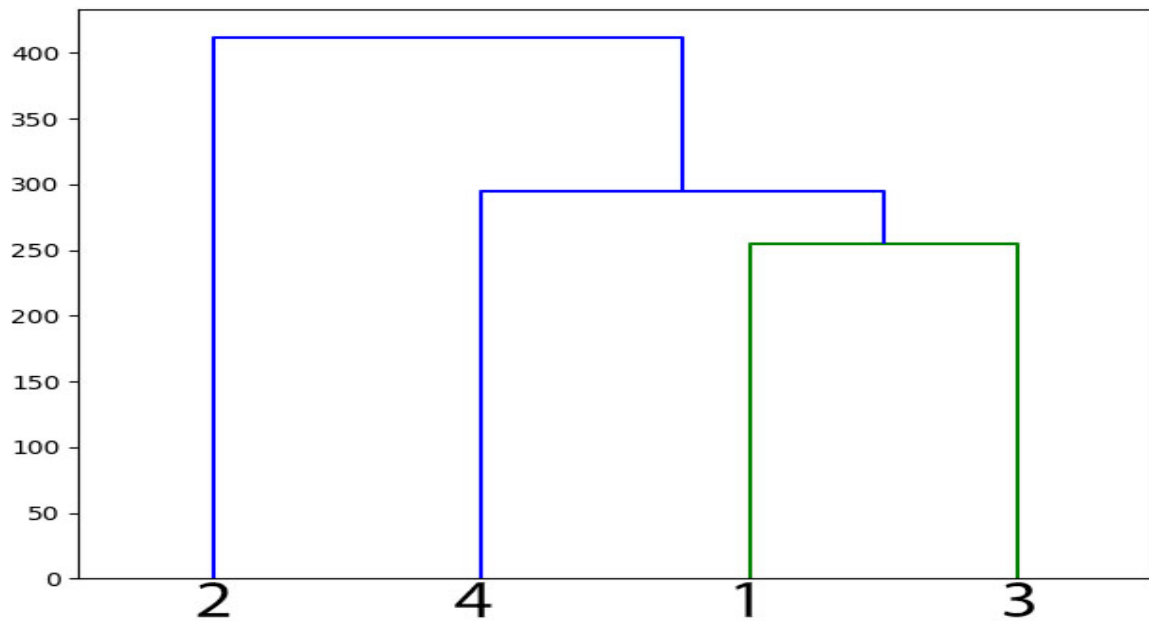
دانشجو 4	دانشجو 2	دانشجو 1 و 3	
2.44	2.82	0	دانشجو 1 و 3
4.24	0	2.82	دانشجو 2
0	4.24	2.44	دانشجو 4

با توجه به اطلاعات بالا فاصله خوشه دانشجوی 1,3 و دانشجوی 4 کمترین فاصله است. پس این دو خوشه باهم ترکیب میشوند. و در جدولی به صورت زیر قرار میگیرند.

دانشجو 2	دانشجو 1 و 3 و 4	
2.82	0	دانشجو 1 و 3 و 4
0	2.82	دانشجو 2

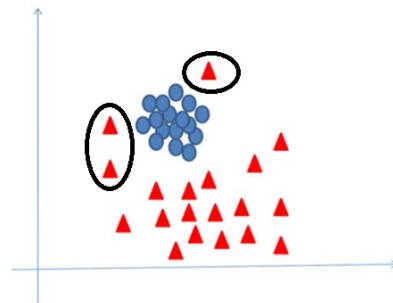
و در انتها نیز خوشه دانشجوی 1 و 3 و 4 با خوشه دانشجوی 2 ادغام میشود و تشکیل یک خوشه را میدهد و عمل خوشه‌بندی تمام میشود.

همچنین نمودار dendrogram خوشه‌بندی بالا به صورت زیر است:



سوال پنجم:

الگوریتم استفاده شده در این سوال نمی‌تواند الگوریتم kmeans باشد؛ زیرا در این صورت داده‌هایی که در شکل زیر با بیضی مشخص شده‌اند باید در خوشه که با علامت دایره مشخص شده است خوشه‌بندی میشدند زیرا به مرکز این خوشه نزدیکتر هستند.



همچنین الگوریتم استفاده شده در این سوال میتواند الگوریتم EM باشد. زیرا به نوعی داده‌ها را بر اساس توزیع آنها خوشه‌بندی کرده است.