

به نام خدا

دانشگاه صنعتی امیرکبیر

دانشکده کامپیوتر و فناوری اطلاعات

داده کاوی

نیمسال اول ۹۷-۹۸

تمرینات سری چهارم

موعد تحویل: ۹۷/۱۰/۲۳

- ۱- با توجه به SVM دو کلاسه شرح داده شده در کلاس، یک سناریو برای SVM چندکلاسه (M کلاس) با ذکر توضیحات ارائه دهید
- ۲- برای تولید کاندیدهای k frequent itemset تایی یک رویکرد مورد استفاده که مطرح شد، روش $F_{k-1} \times F_{k-1}$ بود. اثبات کنید رویکرد شرح داده شده که صرفاً دو $k-1$ itemset تایی را ترکیب می کرد که $k-2$ تای اول آنها یکسان باشند، کامل است. یعنی اثبات کنید همه itemset های k تایی frequent بدین ترتیب تولید خواهند شد.
- ۳- جدول زیر نوع کتاب های خریداری شده توسط مشتریان یک فروشگاه را نشان می دهد.

نوع کتاب خریداری شده	شماره خرید
مذهبی، رمان، شعر	۱
مذهبی، رمان، تاریخی	۲
روانشناسی، رمان، تاریخی، شعر	۳
روانشناسی، تاریخی، شعر	۴
روانشناسی، مذهبی، رمان	۵
روانشناسی، مذهبی، رمان، تاریخی	۶
روانشناسی	۷
روانشناسی، مذهبی، رمان	۸
روانشناسی، مذهبی، تاریخی	۹
روانشناسی، مذهبی	۱۰

الف) ساختار شبکه ای مربوط به جدول بالا را رسم کرده و با در نظر گرفتن حد آستانه support برابر 30% هر نود این ساختار را با برچسب های زیر برچسب زنی نمایید

M : maximal frequent itemset, C : closed frequent itemset, N : frequent but neither maximal nor closed,

• **I** : infrequent.

ب) با استفاده از روش FP-Growth تمامی frequent itemset هایی که به کتاب نوع رمان ختم می شود، بیابید. در این قسمت حدآستانه support را برابر ۲۰٪ در نظر بگیرید.

ج) با در نظر گرفتن حدآستانه confidence برابر ۰.۵ قواعد انجمنی معتبر قابل استخراج از frequent itemset های به دست آمده از قسمت ب را بیابید.

۴- مسئله خوشه بندی دانشجویان بر مبنای ۳ نمره از ۳ درس آنها مدنظر است. چهار دانشجوی جدول زیر را با استفاده از روش خوشه بندی سلسله مراتبی خوشه بندی نمایید. با در نظر گرفتن Dendrogram .min مربوطه را رسم نمایید. از معیار فاصله اقلیدسی استفاده کنید

	درس ۱	درس ۲	درس ۳
دانشجو شماره ۱	۱۵	۲۰	۱۸
دانشجو شماره ۲	۱۴	۱۷	۱۷
دانشجو شماره ۳	۱۶	۱۹	۱۷
دانشجو شماره ۴	۱۸	۱۸	۱۸

۵- نتیجه خوشه بندی داده هایی در فضای دو بعدی به صورت زیر است. آیا الگوریتم مورد استفاده می تواند k-means باشد؟ آیا الگوریتم مورد استفاده می تواند EM باشد

