

نام و نام خانوادگي: وزارت علوم و تحقیقات و فناوری عنوان امتحان: داده کاوی شماره دانشجويي: تاريخ امتحان: 30/10/97

نام استاد: ملكوتي مدت امتحان: 2 ساعت

1. در مجموعه داده زیر با فرض min-sup= %30 و min\_conf= %70 :

|  |  |
| --- | --- |
| A,B,C,E | T1 |
| A,C,D,E | T2 |
| B,C,E | T3 |
| A,C,D,E | T4 |
| C,D,E | T5 |
| A,D,E | T6 |

* با استفاده از الگوریتم Apriori الگوهای پرتکرار را پیدا کنید.
* دو قانون معتبر (با ذکر support و confidence ) برای این مجموعه داده بنویسید.
* یک max-pattern و یک closed pattern برای این مجموعه داده بنویسید.

1. روش نمونه گیری برای بهبود الگوریتم Apriori را توضیح دهید.
2. برای مجموعه داده سوال اول یک FP-tree بکشید.
3. در مجموعه داده زیر محاسبه کنید الگوریتم CART کدام صفت را بعنوان ریشه درخت تصمیم انتخاب خواهد کرد؟

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chills | Runny nose | Headache | Fever | Flu? |
| Y | Y | No | N | Y |
| Y | N | Strong | Y | Y |
| N | Y | Mild | Y | Y |
| Y | Y | Mild | Y | Y |
| N | Y | Strong | Y | Y |
| Y | N | Mild | Y | N |
| N | N | No | N | N |
| N | Y | Strong | N | N |

1. با استفاده از Naïve Bayes Classifier کلاس نمونه داده زیر را تعیین کنید:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chills | Runny nose | Headache | Fever | Flu? |
| Y | N | No | Y | ? |

1. با مثال نشان دهید چرا معیارهای Precision و Recall هر کدام به تنهایی نمی توانند معیار خوبی برای دسته بندی باشند.
2. نقاط ضعف الگوریتم k-means برای خوشه بندی چیست؟
3. دو روش برای پیدا کردن داده های پرت بنویسید.

موفق باشید.