

مقطع کارشناسی ارشد	مهلت: ۱۸۰ دقیقه (بدون تمدید)	تاریخ: ۸۸/۱۱/۱۲
دانشگاه آزاد اسلامی واحد زنجان	شماره دانشجویی:	نام و نام خانوادگی:

"مومن از همه مردم گرفتار تر است زیرا باید به کار دنیا و آخرت هر دو برسد" حضرت علی (ع)

سوال اول: وظایف اصلی داده کاوی را بطور دقیق دسته‌بندی کرده و مباحث اصلی هر دسته را نامبرده و بصورت مختصر و مفید شرح دهید.

سوال دوم: Feature Subset Selection چیست؟ چه رهیافتهایی برای انجام آن روی داده‌ها وجود دارد؟ تفاوت آن با Dimension Reduction چیست؟

سوال سوم: نمونه‌های آموزشی زیر را در نظر بگیرید:

X	y	z	تعداد نمونه‌های کلاس C1	تعداد نمونه‌های کلاس C2
0	0	0	5	40
0	0	1	0	15
0	1	0	10	5
0	1	1	45	0
1	0	0	10	5
1	0	1	25	0
1	1	0	5	20
1	1	1	0	15

الف) یک درخت تصمیم تا دوسطح برای این نمونه‌ها ایجاد کنید. برای Splitting از معیار Gini استفاده نمایید.

ب) در درخت تصمیم بدست آمده Entropy هریک از گره‌های برگ را محاسبه نمایید.

سوال چهارم: مجموعه داده های زیر را در نظر بگیرید:

Tid	Refund	Marital Status	Taxable Income	Evad
1	Yes	Single	125k	No
2	No	Married	100k	No
3	No	Single	70k	No
4	Yes	Married	120k	No
5	No	Divorced	95k	Yes
6	No	Married	60k	No
7	Yes	Divorced	220k	No
8	No	Single	85k	Yes
9	No	Married	75k	No
10	No	Single	90k	Yes

فرض کنید احتمال تعلق هر صفت پیوسته Taxable Income به هر یک از کلاسهای Yes و No دارای توزیع نرمال باشد. همچنین فرض کنید هر یک از صفتها برای هر کلاس مفروض مستقل از هم باشند. همچنین رکورد تست زیر را در نظر بگیرید:

$$X=(\text{Refund}=\text{Yes}, \text{Married}, \text{Income}=80)$$

با استفاده از naïve Bayes Classifier کلاس این نمونه تست را تعیین نمایید.

**سوال پنجم:** با فرض داشتن داده‌های سوال چهارم، با استفاده از الگوریتم<sup>1</sup> PEELS فاصله بین رکورد X و Y را محاسبه نمایید:

Tid	Refund	Marital Status	Taxable Income	Evad
X	Yes	Single	125k	No
Y	No	Married	100k	No

**سوال ششم:** تراکنش‌های سبد خرید جدول زیر را در نظر بگیرید:

Transaction ID	Items Bought
1	{a,b,d,e}
2	{b,c,d}
3	{a,b,d,e}
4	{a,c,d,e}
5	{b,c,d,e}
6	{b,d,e}
7	{c,d}
8	{a,b,c}
9	{a,d,e}
10	{b,d}

(الف) ماکزیمم تعداد قوانین که می‌تواند از این Data Set استخراج شود، چقدر است؟ (شامل قوانینی که support صفر دارند)

(ب) ماکزیمم اندازه itemset تکراری که می‌تواند استخراج شوند، چقدر است؟ (فرض کنید  $\text{minsup} > 0$ )

(ج) یک Itemset (از اندازه ۲ یا بزرگتر) پیدا کنید که بزرگترین support را داشته باشد.

(د) یک زوج از itemها مانند a و b پیدا کنید که قوانین  $\{a\} \rightarrow \{b\}$  و  $\{b\} \rightarrow \{a\}$  دارای confidence یکسانی باشند.

**سوال هفتم:** با استفاده از ماتریس مجاورت زیر نقاط داده شده را طبق الگوریتم Group Average، بطور کامل بصورت سلسله مراتبی کلاستر بندی کرده و درخت dendrogram آن را رسم نمایید. در هر مرحله از الگوریتم ماتریس مجاورت بروزرسانی شده را بدست آورید.

Points	P1	P2	P3	P4	P5	P6
P1	1	0.7	0.35	0.9	0.2	0.5
P2	0.7	1	0.55	0.45	0.1	0.8
P3	0.35	0.55	1	0.6	0.3	0.2
P4	0.9	0.45	0.6	1	0.75	0.4
P5	0.2	0.1	0.3	0.75	1	0.85
P6	0.5	0.8	0.2	0.4	0.85	1

موفق و پیروز باشید - علی امیری