

بسمه تعالی

تکلیف سری اول درس داده کاوی

- 1- فرایند اکتشاف دانش در داده کاوی را توضیح دهید.
- 2- فرض کنید دو آبجکت با ویژگی های $(22,1,42,10)$ و $(20,0,36,8)$ داریم، موارد زیر را محاسبه کنید:
الف) فاصله اقلیدسی، منهتن و مینکوفسکی با $h=3$ بین دو آبجکت را بدست آورید.
ب) فاصله *supremum* بین دو آبجکت را محاسبه کنید.
- 3- فرض کنید دیتاست دو بعدی زیر را داریم:

	A_1	A_2
x_1	1.5	1.7
x_2	2	1.9
x_3	1.6	1.8
x_4	1.2	1.5
x_5	1.5	1.0

- با توجه به یک نقطه داده جدید، $x=(1.4,1.6)$ به عنوان یک پرس و جو، نقاط پایگاه داده را بر اساس شباهت نقطه جدید با استفاده از فاصله اقلیدسی، فاصله منهتن و شباهت کسینوس رتبه بندی کنید.
- 4- مجموعه داده های یک بعدی $x=\{5,23,17.6,7.32,1.11\}$ مفروض است. با استفاده از روش تغییر شکل داده های *min-max* در فاصله $[-1,1]$ آنها را نرمال سازی کنید.
 - 5- مجموعه داده زیر را داریم. با استفاده از روش *Binning* آنها را در سه طبقه قرار داده و بر اساس میانگین و مرزهای پایینی و بالایی در هر طبقه آنها را هموارسازی کنید.

3 – 7 – 14 – 18 – 23 – 25 31 – 42 – 47

- 6- پایگاه کتابفروشی ما به صورت زیر است و *min-support* 50% می باشد. آیتم های پرتکرار را با استفاده از الگوریتم *Eclat* بیابید.

ITEMS	
Jane Austen	A
Agatha Christie	C
Sir Arthur Conan Doyle	D
Mark Twain	T
P. G. Wodehouse	W

DATABASE	
Transaction	Items
1	A C T W
2	C D W
3	A C T W
4	A C D W
5	A C D T W
6	C D T

- 7- اگر یک مسئله دسته بندی داشته باشید، چه روشی را برای کاهش بعد پیشنهاد می دهید؟ چرا؟
- 8- تراکنش های سبد خرید جدول زیر را در نظر بگیرید:

الف) الگوهای مکرر را با استفاده از الگوریتم آپریوری بدست آورید. $Minsub=2$

ب) قوانین قوی را برای آن استخراج کنید. $Minconf = 70\%$

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}

9- یک دیتابیس با پنج Transaction به صورت زیر داریم. با فرض $60\% = min_sup$ و $80\% = min_conf$

الف) تمام الگوهای مکرر را با استفاده از الگوریتم آپریوری بدست آورید.

ب) تمامی قوانین انجمنی قوی را استخراج کنید.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}