

بسمه تعالی

تکالیف سری اول درس یادگیری ماشین

- 1- فرآیند اکتشاف دانش در داده کاوی را توضیح دهید.
- 2- فرض کنید دو ویژگی های $(22,1,42,10)$ و $(20,0,36,8)$ را داریم،
الف) فاصله اقلیدسی، منهتن و مینکوفسکی با $h=3$ یا 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 پایانی و بالایی در هر طبقه، آنها را هموارسازی کنید.

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

- 6- در یک سیستم پیش بینی ترک شغل، دیتابیس زیر را داریم:

SalaryLevel	Overtime	Quit
Low	Yes	Yes
Medium	Yes	Yes
High	No	No
Low	No	No
Medium	Yes	No
Medium	No	No
High	Yes	Yes
Low	Yes	Yes
Medium	No	No

با استفاده از Naive Bayes تعیین کنید برای کارمند جدید با SalaryLevel = Medium و Overtime = Yes آیا احتمال ترک شغل (Quit=Yes) بیشتر است یا نه؟

7- فرض کنید یک مجموعه داده آموزشی به شکل زیر دارید که شامل ویژگی‌های دو بعدی و برچسب‌های کلاس است، حال یک داده جدید با ویژگی‌های $X_1=4$ و $X_2=4$ وارد می‌شود و شما می‌خواهید پیش‌بینی کنید که این داده به کدام کلاس تعلق دارد؟ الگوریتم KNN را با $K=3$ اجرا کنید.

ویژگی 1 (X_1)	ویژگی 2 (X_2)	کلاس
2	3	A
3	3	A
6	5	B
7	6	B
8	7	B

8- داده‌های زیر مربوط به دو ویژگی با مقیاس متفاوت است:

نمونه	قد (cm)	وزن (kg)	کلاس
A	160	60	C_1
B	180	80	C_2
C	170	72	?

می‌خواهیم کلاس نمونه‌ی C را با KNN و $K=1$ تعیین کنیم. اگر از فاصله‌ی اقلیدسی استفاده کنیم، (الف) بدون نرمال‌سازی چه کلاسی پیش‌بینی می‌شود؟ (ب) با نرمال‌سازی Min-Max در بازه $[0,1]$ چه تغییری در نتیجه ایجاد می‌شود؟