

به نام خدا



دانشگاه تهران

دانشکده فنی

دانشکده مهندسی برق و کامپیوتر



درس داده کاوی

تمرین ۱

اسفند ماه ۱۴۰۰

* فهرست

سؤال ۱.....	۳
سؤال ۲.....	۴
سؤال ۳.....	۵
ملاحظات (حتما مطالعه شود).....	۶

سؤال ۱

الف) برای هر کدام از موارد زیر ابتدا مشخص کنید که خصیصه گسسته، پیوسته یا باینری است و سپس، نوع آن شامل Nominal، ordinal، Interval-scaled و Ratio-scaled را مشخص فرمایید.

- Continues سرعت خودروها بر حسب mph Ratio
- Discrete شدت بارش که با مقادیر: بدون بارش، بارش متناوب، بارش بی وقفه مشخص می‌شود. Ordinal
- Discrete تجزیه نور به رنگ‌های رنگین کمان Ordinal or Nominal
- Discrete اندازه‌گیری هوش افراد با استفاده از تست IQ Ratio
- Continuous عدد بارکد اجناس در یک سوپرمارکت Ordinal

ب) برای هر کدام از خصیصه‌ها در جدول زیر، نمودارهای مناسب جهت مصورسازی داده را مشخص نمایید.

	Nominal	Ordinal	Interval-scaled	Ratio-scaled
Histogram		*		
Pie chart				
Box-Plot				
Bar Chart	*			

سؤال ۲

دو خصیصه A و B را در نظر بگیرید. برای هر کدام ۲۰ مقدار جمع‌آوری شده است که در ادامه قابل مشاهده است. با توجه به مقادیر به سوالات زیر پاسخ دهید.

A: 11.7, 75.3, 6.21, 265.6, 0.31, 43.02, 111.87, 145, 121.2, 95, 111, 42.5, 0.04, 72, 58.1, 659.98, 110.02, 0.78, 598.23, 311.54

B: 17.20, 12.90, 0.53, 34.80, 4.20, 1.50, 3.90, 6.70, 7.43, 8.60, 11.30, 12.10, 5.6, 13.20, 15.80, 19.80, 23.90, 34.17, 2.30, 7.80

الف) میانگین، چارک اول، چارک سوم و انحراف معیار را برای هر کدام از خصیصه‌ها محاسبه کنید.

ب) نمودار جعبه‌ای برای دو خصیصه را رسم کنید و میزان پراکندگی داده‌ها در این دو خصیصه را مقایسه کنید.

ج) نمودار پراکندگی برای دو خصیصه را رسم کنید.

د) توزیع مقادیر دو خصیصه A و B را با رسم نمودار Q-Q plot مقایسه کنید.

ه) آیا این دو خصیصه با هم هم‌بستگی^۱ دارند؟ توضیح دهید.

سؤال ۳

مجموعه کلمات X_1 تا X_5 را به صورت بردار دو بعدی با مقادیر زیر مدل کرده‌ایم.

	A1	A2
x1	1.5	1.7
x2	2	1.9
x3	1.6	1.8
x4	1.2	1.5
x5	1.5	1.0

الف) مجموعه کلمات X_1 تا X_5 را بر اساس مرتبط‌ترین کلمه با کلمه پرسش $X=(1.4,1.6)$ با استفاده از معیارهای اقلیدسی، منهتن، سوپریمم و فاصله کسینوسی رتبه‌بندی کنید.

ب) مقادیر خصیصه‌ها را با استفاده از z -score نرمال کنید و سپس مجموعه کلمات را با استفاده از فاصله اقلیدسی نسبت به کلمه پرسش رتبه‌بندی کنید.

ج) توضیح دهید چگونه می‌توان عدم شباهت میان خصیصه‌های ratio-scaled را محاسبه کرد؟

ملاحظات (حتما مطالعه شود)

گزارش شما باید در قالب یک فایل با فرمت PDF (گزارش تایپ شده) و با عنوان DM_HW1_StudentID تحویل داده شود.

- خوانایی گزارش شما از اهمیت ویژه‌ای برخوردار است. به تمرین‌هایی که به صورت کاغذی تحویل داده شوند یا به صورت عکس در سایت بارگذاری شوند، ترتیب اثری داده نخواهد شد.
- مهلت تحویل تمرین به هیچ عنوان تمدید نخواهد شد. تمرین تا یک هفته بعد از مهلت تعیین شده با جریمه تحویل گرفته می‌شود که جریمه تاخیر تحویل تمرین تا یک هفته ۳۰ درصد است.
- توجه کنید این تمرین باید به صورت تک نفره انجام شود و پاسخ‌های ارئه شده باید نتیجه فعالیت فرد نویسنده باشد (همفکری و به اتفاق هم نوشتن تمرین نیز ممنوع است). در صورت مشاهده تقلب به همه افراد مشارکت کننده، نمره تمرین صفر و به استاد نیز گزارش می‌گردد.
- در صورت بروز هرگونه مشکل با ایمیل زیر در ارتباط باشید:

<mailto:bazmi.parisa@gmail.com>

مهلت تحویل بدون جریمه: جمعه ۱۳ اسفند ۱۴۰۰

مهلت تحویل با تاخیر، با جریمه ۳۰ درصد: جمعه ۲۰ اسفند ۱۴۰۰