

نام و نام خانوادگي: وزارت علوم و تحقیقات و فناوری عنوان امتحان: داده کاوی(میان ترم مجدد) شماره دانشجويي: تاريخ امتحان: 28/3/97

نام استاد: ملكوتي مدت امتحان: 60 دقیقه

1. میانگین هرس شده و میانگین وزنی را در اندازه گیری مرکزیت داده ها توضیح دهید.
2. نمودار چندک یا Quantile را برای داده های زیر رسم کنید:

25 27 32 50 55 75 81 122 125 140

1. فرض کنید اطلاعات سه دانشجو به شرح زیر در پایگاه داده موجود است. مقطع تحصیلی صفت ترتیبی حاوی چهار مقدار کاردانی، کارشناسی، کارشناسی ارشد و دکتری است. رشته تحصیلی صفت اسمی و معدل صفت عددی در بازه 0 تا 20 می باشد. ماتریس عدم شباهت را برای این نمونه ها محاسبه کنید:

|  |  |  |  |
| --- | --- | --- | --- |
| معدل(صفت عددی) | رشته تحصیلی(صفت اسمی) | مقطع(صفت ترتیبی) | شماره |
| 18 | مهندسی کامپیوتر | کارشناسی ارشد | 1 |
| 15 | مهندسی عمران | کارشناسی | 2 |
| 20 | مهندسی کامپیوتر | دکتری | 3 |

1. به اختصار بنویسید مهمترین مسائل و مشکلات در تجمیع داده ها از منابع مختلف کدامند و چه راه حل هایی دارند.
2. کاربرد تبدیل موجک در پیش پردازش داده ها چیست؟
3. شکل معماری چند سطحی انبار داده را بطور کامل ترسیم کنید.
4. شکل زیر یک نمونه مکعب داده تعریف شده در شهرداری تهران است. ضمن مشخص کردن ویژگی های بعد (dimention) و معیار(measure) در این مکعب، مدل مفهومی آن را با استفاده از شمای ستاره ای ترسیم کنید:

