1. جدول زیر تراکنشهای یک فروشگاه را نشان می‌دهد. با فرض اینکه minsup>0 است به پرسشهای زیر پاسخ دهید.

الف) بیشترین تعداد قوانین انجمنی (با پشتیبان صفر یا بیشتر) که می‌توان از این داده‌ها استخراج نمود چند تا است؟

ب) بیشترین طول اقلام تکرار شونده که می‌توان از این داده‌ها استخراج نمود چند تا است؟

|{Milk, Diapers, Bread, Butter}| = 4

ج) بیشترین تعداد اقلام تکرار شونده با طول 3 که می‌توان از این داده‌ها استخراج نمود چند تا است؟

{Milk, Beer, Diapers}

{Bread,Butter,Milk}

{Milk, Diapers,Cookies}

{Bread,Butter,Cookies}

{Beer,Cookies,Diapers}

{Milk,Diapers,Bread}

{Milk,Diapers,Butter}

{Diapers,Bread,Butter}

= 8

1. مجموعه داده‌های زیر داده شده است که A و B و C سه متغیر تصادفی دودویی و y نیز یک متغیر دودویی است. هدف این است که از روی متغیرهای A و B و C متغیر Y را پیش بینی نماییم.
2. الف) اگر نمونه A=0, B=0, C=1 باشد الگوریتم بیز ساده کدام دسته را پیش بینی می‌کند؟

x=(0,0,1)

p(y=0|x) = p(x|y=0)p(y=0) , p(y=1|x) = p(x|y=1)p(y=1)

p(y=0) = 3/7, p(y=1) = 4/7

p(A=0, B=0, C=1|y=0) = p(A=0|y=0)\*p(B=0|y=0)\*p(C=1|y=0)= 2/3\*1/3\*1/3=2/27

p(A=0,B=0,C=1|y=1)=p(A=0|y=1)\*p(B=0|y=1)\*p(C=1|y=1)=1/4\*2/4\*2/4=1/16

p(y=0|x)=2/27\*3/7=2/63 = 8/(4\*7\*9)

p(y=1|x)=1/16\*4/7=1/28 = 9/(4\*7\*9)

🡺 y=1

1. ب) اگر A, B, C سه متغیر تصادفی مستقل از هم باشند آیا دسته بند دیگری می‌توان داشت که از بیز ساده بهتر عمل نماید؟ توجه کنید که برای این بخش مجموعه داده زیر مهم نیست.

بله، از استقلال نمی‌توان استقلال شرطی را نتیجه گرفت (چون Y از A, B, C مستقل نیست) پس دسته بند بیز ساده بهترین نیست.

1. به کتاب مراجعه کنید.
2. به کتاب مراجعه کنید.
3. مجموعه داده‌های زیر داده شده است.

الف) با استفاده از روش هموارسازی با میانگین سبد با اندازه 3، داده های بالا را هموار نمایید.

راه حل: به این کار binning می‌گویند و از روش‌های گسسته سازی است. هر سه عدد را در یک سبد می‌گذاریم و آنها را با میانگین آن سبد جایگزین می‌کنیم.

ب) از روش نرمال سازی Min-Max، داده 35 (در جدول بالا) را به بازه [0,1] تبدیل نمایید.