

دوره آمار و ریاضیات تحلیل گر داده

آکادمی همراه اول

مقدمه

در ادامه یک پژوهه جامع و کاربردی را معرفی می‌کنیم که در آن بسیاری از مفاهیم تدریس شده در طول کلاس به کار می‌رود. در این پژوهه از دیتاست واقعی تبلیغات^۱ استفاده می‌کنیم که در آن هزینه‌های تبلیغاتی در رسانه‌های تلویزیون، رادیو و روزنامه به همراه فروش ثبت شده است. علاوه بر تحلیل توصیفی و گرافیکی داده‌ها، از عملیات ماتریسی، کاهش ابعاد بر پایه تجزیه مقدار منفرد^۲ و حل سیستم معادلات خطی برای برآورد احتمال رخداد «فروش بالا» در این پژوهه گنجانده می‌کنیم. همچنین آزمون‌های آماری، استدلال بیزی برای برآورد احتمال رخداد «فروش بالا» در این پژوهه گنجانده شده‌اند.

طرح کلی پژوهه

(۱) کشف و پردازش داده^۳: با استفاده از آمار توصیفی (میانگین، میانه، واریانس) و نمودارهایی مانند هیستوگرام و نمودار جعبه‌ای، ساختار داده و توزیع هر یک از متغیرها (Sales, Newspaper, Radio, TV) را بررسی می‌کنیم.

(۲) تحلیل ماتریس و کاهش ابعاد: محاسبه ماتریس کوواریانس هزینه‌های تبلیغات، استخراج مقادیر ویژه و بردارهای ویژه، انجام تجزیه SVD بر روی داده‌های نرمال شده و تفسیر نتایج (نمایش نسبت واریانس توضیح داده شده توسط هر جزء اصلی)

(۳) مدل‌سازی رگرسیون خطی: تشکیل یک مدل خطی ($Sales = \beta_0 + \beta_1 \times TV + \beta_2 \times Radio + \beta_3 \times Newspaper$)، برآش مدل با استفاده از روش حداقل مربعات به وسیله حل سیستم معادلات خطی (استفاده از معکوس ماتریس یا شبیه معکوس به کمک SVD)، استخراج نتایج آماری شامل ضرایب مدل، آزمون t و p-value

¹ Advertising

² SVD

³ Exploratory Data Analysis

دوره آمار و ریاضیات تحلیل‌گر داده

آکادمی همراه اول

4) استدلال بیزی: تعریف «فروش بالا» به عنوان فروش‌هایی که بالاتر از میانه فروش هستند، محاسبه تعداد نمونه‌های فروش بالا، به روزرسانی یک توزیع پیشین (توزیع Beta با پارامترهای اولیه ۱ و ۱) بر اساس داده به دست آمده و برآورد میانگین پسین احتمال رخداد فروش بالا