

## مقدمه

در ادامه یک پروژه جامع و کاربردی را معرفی می‌کنیم که در آن بسیاری از مفاهیم تدریس شده در طول کلاس به کار می‌رود. در این پروژه از دیتاست واقعی تبلیغات<sup>1</sup> 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<sup>2</sup> و حل سیستم معادلات خطی برای برازش مدل رگرسیون استفاده می‌کنیم. همچنین آزمون‌های آماری، استدلال بیزی برای برآورد احتمال رخداد «فروش بالا» در این پروژه گنجانده شده‌اند.

## طرح کلی پروژه

(1) کشف و پردازش داده<sup>3</sup>: 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.

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

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

---

<sup>1</sup> Advertising

<sup>2</sup> SVD

<sup>3</sup> Exploratory Data Analysis

## دوره آمار و ریاضیات تحلیل گر داده

آکادمی همراه اول

4) استدلال بیزی: تعریف «فروش بالا» به عنوان فروش هایی که بالاتر از میانه فروش هستند، محاسبه تعداد نمونه های 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