Checkpoint 2

$$D = \begin{pmatrix} 4 & 8 & 12 \\ 6 & 10 & 15 \\ 5 & 7 & 9 \\ 3 & 6 & 9 \end{pmatrix}$$

Row = Different region

Column = Specific product category (A, B, C)

Tasks

① Identify the data type & matrix elements

Type integer (or numeric)

② Calculate the ~~Mean~~ mean, median, std for each column (Pdt A, B, & C)

- ~~Mean~~ Product A (First column, 4, 6, 5, 3)

$$\text{Mean} = \frac{4+6+5+3}{4} = 4.5$$

- Median:

Sorted data = 3, 4, 5, 6

$$\text{Median} = \frac{4+5}{2} = 4.5$$

- Standard deviation

$$\text{Variance} = \frac{(4-4.5)^2 + (6-4.5)^2 + (5-4.5)^2 + (3-4.5)^2}{4}$$

$$= \frac{0.25 + 2.25 + 0.25 + 2.25}{4} = \frac{5}{4} = 1.25$$

Std = $\sqrt{\text{variance}}$

$= \sqrt{1.25}$

$= 1.118$ //

* Product B (second column, 8, 10, 7, 6)

- Mean

$= \frac{8+10+7+6}{4}$

$= \frac{31}{4} = 7.75$ //

- Median = ~~$\frac{10+7}{2}$~~ 6, 7, 8, 10

$\Rightarrow \frac{7+8}{2} = 7.5$ //

- Standard deviation

Variance = $\frac{\sum (x-\bar{x})^2}{n}$

$= \frac{(8-7.5)^2+(10-7.5)^2+(7-7.5)^2+(6-7.5)^2}{4}$

$= \frac{0.0625+5.0625+0.5625+3.0625}{4} = \frac{8.75}{4} = 2.1875$

Std = $\sqrt{\text{variance}}$

$= \sqrt{2.1875} = 1.479$

Product C (Third column, 12, 15, 9, 9)

Mean = $\frac{12+15+9+9}{4}$

$= \frac{45}{4}$

$4 = 11.25$

Median

Sorted data = 9, 9, 12, 15

$\Rightarrow \frac{9+12}{2} = 10.5$

Standard deviation

Variance = $\frac{(12-11.25)^2+(15-11.25)^2+(9-11.25)^2+(9-11.25)^2}{4}$

$= \frac{0.5625+14.0625+5.0625+5.0625}{4}$

$= \frac{24.75}{4} = 6.1875$

Std = $\sqrt{6.1875}$

$= 2.487$

⑥ For each column determine the range and variance

* Product A (First column, 4, 6, 5, 3)

- Range = Max − Min

6 − 3 = 3//

- Variance

Mean = 4.5,

Variance = 1.25

* Product B (2nd column 8, 10, 7, 6)

Range = Max − Min

= 10 − 6 = 4//

Variance = 2.1875//

* Product C (3rd column 12, 15, 9, 9)

Range = Max − Min

= 15 − 9 = 6//

Variance = 6.1875,

Part 2: Linear Algebra

1 Compute the transpose of a matrix

$$D = \begin{pmatrix} 4 & 6 & 5 & 3 \\ 8 & 10 & 7 & 6 \\ 12 & 15 & 9 & 9 \end{pmatrix}$$

2) Multiply the matrix D by a price vector P where: $P = [5 \ 10 \ 15]$

Solution

- First row

$(4 \times 5) + (8 \times 10) + (12 \times 15)$

$20 + 80 + 180 = 280$,

- Second row

$(6 \times 5) + (10 \times 10) + (15 \times 15)$

$30 + 100 + 225 = 355$,

- Third row

$(5 \times 5) + (7 \times 10) + (9 \times 15)$

$25 + 70 + 135 = 230$,

- Fourth row

$(3 \times 5) + (6 \times 10) + (9 \times 15)$

$15 + 60 + 135 = 210$,

$\therefore D * P = \begin{pmatrix} 280 \\ 355 \\ 230 \\ 210 \end{pmatrix}$

* This vector represents the total sales value for each region based on ē given price vector P.

3) Determine which region generates the highest revenue
Region 2 with a total revenue of 355

- For each product, calculate the total revenue across all regions
* Product A (1st column 4, 6, 5, 3)
~~(4x5)+(6x5~~

$$D = \begin{pmatrix} 4 & 8 & 12 \\ 6 & 10 & 15 \\ 5 & 7 & 9 \\ 3 & 6 & 9 \end{pmatrix}$$

$$P = \begin{pmatrix} 5 \\ 10 \\ 15 \end{pmatrix}$$

- Product A (1st column 4, 6, 5, 3)
(4×5) + (6×5) + (5×5) + (3×5)
20 + 30 + 25 + 15 = 90//

- Product B (2nd column 8, 10, 7, 6)
(8×10) + (10×10) + (7×10) + (6×10)
80 + 100 + 70 + 60 = 310//

- Product C (3rd column 12, 15, 9, 9)
(12×15) + (15×15) + (9×15) + (9×15)
180 + 225 + 135 + 135 = 675

Thus the ~~overt~~ total revenue across all regions for each product is
Product A: 90
Product B: 310
Product C: 675
~~Pro~~

~~H~~ What does the mean tell you about the typical sales for each product
- Product A
Mean = 4.5
This means that, on average each region

How does the coefficient of variation help you understand sales consistency across products?

- The coeffient of variation (CV) is a standardized measure of dispersion that is calculated as ē ratio ō ē standard deviation to ē mean, often expressed as a percentage. It helps you understand ē relative variability of sales for each pdt, ẅh is useful for comparing consistency across pdts ẅ different means.

Formula

$$CV = \left(\frac{\text{Standard deviation}}{\text{Mean}}\right) \times 100\%$$

- A lower CV indicates more consistent sales (less variability relative to the mean)
- A higher CV indicates less consistent salesss (more variability relative to ē mean)

Product A

Mean = 4.5

Std = ~~1.479~~ 1.118

~~CV = 1.479~~ $CV = \left(\frac{1.118}{4.5}\right) \times 100\% = 24.84\%$

Product B

Mean = 7.75

Std = 1.479

$CV = \left(\frac{1.479}{7.75}\right) \times 100\% = 19.08\%$

Product C

Mean = 11.25

Std = 2.487

$CV = \left(\frac{2.487}{11.25}\right) \times 100\% = 22.11\%$

5) Discuss how matrix multiplication is used in real-world data science problems (eg revenue calculations, recommendation systems)

Answer

Matrix multiplication is a fundamental operation in data science and is widely used in various real-world applications.

1. Revenue calculations

Matrix multiplication is often used to calculate total revenue by multipling a sales matrix by a price vector. This approach is efficient for handling large datasets and can be easily extended to include additional dimensions such as discounts or taxes

2. Natural Language Processing (NLP)

Matrix multiplication is used in various NLP tasks such as word embeddings, neural networks. This helps in capturing semantic relationships btw words and improves performance of NLP models.

3. Image Processing
In image processing, matrix multiplication is used in operations such as filtering and transformations. For example convolution (apply filters (kernels) to images for tasks like edge detection or blurring), affine transformations (perform operations like rotation, scaling and translation on images). These operations are essential for tasks like computer vision and image recognition.

4. Recommendation Systems
Matrix multiplication is crucial in collaborative filtering, a common technique in recommendation systems. e.g user-item matrix (rows represent users, columns represent items and entries represent ratings or interactions), matrix factorization, predicted ratings. This method helps in predicting user preferences and suggests relevant pdts or content.

5. Network analysis

In network analysis, matrix multiplication is used to study relationships and connections. E.g adjacency matrix (represent connections btw nodes in a network), power of matrix (compute the number of paths of a certain length btw nodes). This is useful for analysing social networks, transportation systems, and other complex networks.

* Matrix multiplication is a versatile tool in data science, enabling efficient computations and transformations across various domains. Its applications range from simple revenue calculations to complex tasks like recommendation systems, image processing and network analysis. Understanding and leveraging matrix multiplication can significantly enhance one's ability to analyse and interpret data in real-world scenarios.