

سوال اول:

1) خوشه بندی بیماران براساس پارامترهایی مانند نوع بیماری، جنسیت و سن (clustering)

هدف: خوشه بندی مراجعان برای ارائه خدمات مناسب به هر دسته

2) پیش بینی امکان ابتلای یک فرد مراجع به بیماری های دیگر (classification)

هدف: شناسایی بیماری های بالقوه در فرد برای جلوگیری از بروز و ابتلای آینده.

3) شناسایی بیماری های ارثی در یک خانواده براساس نام فرد و بیماری (clustering)

هدف: آگاهی از بیماری های احتمالی فرد در آینده یا حال حاضر

4) پیش بینی بیماری هایی که یک فرد در محیط کارش میتواند با آن مواجه شود (classification)

هدف: اتخاذ روش های مناسب در محیط کار فرد برای حداقل کردن آسیب.

سوال 2:

نکته مهم: در صورتی که مورد انتخابی شما غیر از جواب سوال باشد (همانطور که در حل تمرین هم گفته شد)، تنها با استدلال دقیق و منطقی جواب پذیرفته خواهد شد.

● سن بر حسب سال

Not binary, Discrete, Nominal

● روشنایی که با نورسنج اندازه گیری می شود

Continuous, quantitative, ratio

● روشنایی که با نظر افراد بیان می شود

Discrete, qualitative, ordinal

● زاویه اندازه گیری شده با وسیله اندازه گیری (نقاله و ...)

Continuous, quantitative, ratio

● مدال های اهدایی در مسابقات المپیک

Discrete, qualitative, ordinal

● ارتفاع از سطح دریا

Continuous, quantitative, interval/ratio

● تعداد بیماران یک بیمارستان

Discrete, quantitative, ratio

● شماره ISBN

Discrete, qualitative, nominal

سوال 3:

میانگین:

$$\text{Mean Stock Price} = (10+7+20+12+75+15+9+18+4+12+8+14) / 12 = 204 / 12 = \$17$$

میانها:

اعداد را سورت میکنیم. اگر تعداد اعداد زوج بود که میانگین دو عدد وسط برابر میانها خواهند شد. اگر فرد بود خود عدد وسط برابر میانها است.

$$\text{Median Stock Price} = \text{mean of center values } \{4,7,8,9,10,12,12,14,15,18,20,75\} = 24/2 = \$12.$$

مد:

$$\text{Mode Stock Price} = \text{highest frequency of } \{4,7,8,9,10,12,12,14,15,18,20,75\} = \$12$$

انحراف معیار:

انحراف استاندارد نشان دهنده فاصله مورد انتظار یک نقطه انتخاب شده به طور تصادفی از یک مجموعه داده تا مرکز آن مجموعه است و با گرفتن جذر واریانس محاسبه می شود. واریانس میانگین مجموع مجذور فاصله هر نقطه از میانگین مجموعه داده است. لذا واریانس مجموعه داده شده طبق زیر بدست خواهد آمد.

$$\text{Stock Price Variance (Var)} =$$

$$(4-17)^2 + (7-17)^2 + (8-17)^2 + (9-17)^2 + (10-17)^2 + (12-17)^2 + (12-17)^2 + (14-17)^2 + (15-17)^2 + (18-17)^2 + (20-17)^2 + (75-17)^2 =$$

$$(-13)^2 + (-10)^2 + (-9)^2 + (-8)^2 + (-7)^2 + (-5)^2 + (-5)^2 + (-3)^2 + (-2)^2 + (1)^2 + (3)^2 + (58)^2 =$$

$$169 + 100 + 81 + 64 + 49 + 25 + 25 + 9 + 4 + 1 + 9 + 3364 = 3900 / 12 = 325 \$^2.$$

انحراف معیار را از جذر واریانس به صورت زیر بدست می آوریم.

$$\text{Stock Price Standard Deviation (SD) of Stock Price} = \sqrt{(325)} = \pm \$18.03$$

شاخص zscore

$$\text{Z-Score}(X) = [X_i - \text{Mean}(X)] / |\text{SD}(X)|$$

برای تک تک داده ها فرمول بالا اعمال خواهد شد و zscore متناسب با آن در خواهد آمد. نمونه ای از آن را در شکل زیر میتوانید مشاهده کنید. در اینجا شاخص مورد نظر برای داده با مقدار 20 بدست آمده است. به همین ترتیب برای سایر داده ها شاخص بدست خواهد آمد.

$$\text{Z-Score}(\$20) = (\$20 - \$17) / \$18.03 = (\$3) / \$18.03 = 0.1664$$

سوال 4 اختیاری:

Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

- (a) Compute the *Euclidean distance* between the two objects.
- (b) Compute the *Manhattan distance* between the two objects.
- (c) Compute the *Minkowski distance* between the two objects, using $h = 3$.
- (d) Compute the *supremum distance* between the two objects.

Answer:

- (a) Compute the *Euclidean distance* between the two objects.

The Euclidean distance is computed using Equation (2.6).

Therefore, we have $\sqrt{(22 - 20)^2 + (1 - 0)^2 + (42 - 36)^2 + (10 - 8)^2} = \sqrt{45} = 6.7082$.

- (b) Compute the *Manhattan distance* between the two objects.

The Manhattan distance is computed using Equation (2.7). Therefore, we have $|22 - 20| + |1 - 0| + |42 - 36| + |10 - 8| = 11$.

- (c) Compute the *Minkowski distance* between the two objects, using $h = 3$.

The Minkowski distance is

$$d(i, j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \dots + |x_{ip} - x_{jp}|^h} \quad (2.10)$$

where h is a real number such that $h \geq 1$.

Therefore, with $h = 3$, we have $\sqrt[3]{|22 - 20|^3 + |1 - 0|^3 + |42 - 36|^3 + |10 - 8|^3} = \sqrt[3]{233} = 6.1534$.

- (d) Compute the *supremum distance* between the two objects.

The supremum distance is computed using Equation (2.8). Therefore, we have a supremum distance of 6.

The world is not in your books and maps. It is out there.

-Grey gandalf