1. *Topic: fundamental data analysis, software tools*

A mystery data set in file x.csv has 2048 data items (rows), each having 32 real-valued variables (columns). The first row in the file gives the names of the variables.

Write a small program in R or Python that

- loads the data set in x.csv,
- finds the two variables having the largest variances and
- makes a scatterplot of the data items using these two variables.

Your program must read the dataset file from the file and then produce the plot without user intervention.

Your program must work correctly for any dataset file of a similar format. For example, if you permute the rows or columns of the data file, you should get the same output because ordering rows or columns should not affect the result.

Attach a printout of your program code and the scatterplot it produces as an answer to this problem.

Hints: You should see two letters in the scatterplot from which it should be evident that you did okay. If you see something else, then something is wrong.

Solution:

Code:

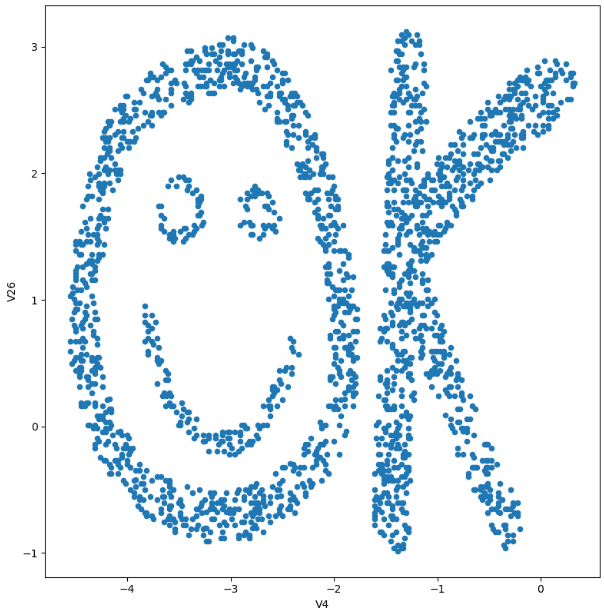
```python
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd


data = pd.read_csv("./x.csv")
variances = data.select_dtypes(include=[np.number]).var()
if type(variances) is not pd.Series or len(variances) < 2:
    raise ValueError("Invalid data")

two_vars = variances.sort_values(ascending=False).take([0,
    1]).keys().to_list()

data.plot(x=two_vars[0], y=two_vars[1], kind="scatter")
plt.show()
```

Scatterplot:



2. *Topic: matrix calculus*

Let A be a 2×2 matrix given by $A_{11} = 1$, $A_{12} = A_{21} = 2$, and $A_{22} = 1.618$.

For the matrix A:

(a) Solve numerically and report the eigenvalues λ_i and column eigenvectors x_i, where $i \in \{1, 2\}$. Normalise the eigenvectors to unit length (if necessary).

(b) Verify that the eigenvectors are orthonormal.

(c) Show, by performing the numerical matrix computation, that A satisfies the equation $A = \sum_{i=1}^{2} \lambda_i x_i x_i^T$.

Hints

You can use R or Python to find eigenvectors and eigenvalues and do matrix and vector multiplications.

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric square matrix. $\lambda \in \mathbb{R}$ is an eigenvalue of A and a column vector $x \in \mathbb{R}^n$ is the corresponding eigenvector if $Ax = \lambda x$. Assume A has n orthonormal eigenvectors $x_i \in \mathbb{R}^n$ and corresponding eigenvalues $\lambda_i \in \mathbb{R}$, where $i \in \{1, \ldots, n\}$. The fact that the eigenvectors are orthonormal means that $x_i^T x_i = 1$ and $x_i^T x_j = 0$ if $i \neq j$.

Solution:

(a) λ is an eigenvalue of A if and only if $\det(A - \lambda I_2) = 0$.

Let's first expand $\det(A - \lambda I_2)$.

$$\begin{aligned}
\det(A - \lambda I_2) &= \det\left(\begin{bmatrix} 1 & 2 \\ 2 & 1.618 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1 & 2 \\ 2 & 1.618 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1-\lambda & 2 \\ 2 & 1.618-\lambda \end{bmatrix}\right) \\
&= (1-\lambda)(1.618-\lambda) - 2 \cdot 2 = \lambda^2 - 2.618\lambda - 2.382
\end{aligned}$$

Since

$$\det(A - \lambda I_2) = 0 \iff \lambda^2 - 2.618\lambda - 2.382 = 0,$$

we can use the quadratic formula to solve λ.

$$\begin{aligned}
\lambda &= \frac{-(-2.618) \pm \sqrt{(-2.618)^2 - 4 \cdot 1 \cdot (-2.382)}}{2 \cdot 1} \\
&= \frac{2.618 \pm \sqrt{16.385924}}{2} = \frac{2.618 \pm 4.0479\ldots}{2}
\end{aligned}$$

Therefore $\lambda_1 \approx 3.333$ and $\lambda_2 \approx -0.715$.
Now lets solve the eigenvector x_1 for the eigenvalue $\lambda_1 \approx 3.333$.
Since $(A - \lambda_i I_2)x_i = 0$,

$$\begin{aligned}
\begin{bmatrix} 1-3.333 & 2 \\ 2 & 1.618-3.333 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\iff \begin{bmatrix} -2.333x_{11} + 2x_{12} \\ 2x_{11} - 1.715x_{12} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

This gives us the system:

$$\begin{cases} -2.333x_{11} + 2x_{12} &= 0 \\ 2x_{11} - 1.715x_{12} &= 0 \end{cases}$$

From the first equation:

$$x_{12} = \frac{2.333x_{11}}{2} = 1.1665x_{11}$$

Setting $x_{11} = 1$, we get $x_{12} = 1.1665$.
So the unnormalized eigenvector is:

$$x_1 = \begin{bmatrix} 1 \\ 1.1665 \end{bmatrix}$$

To normalize to unit length:

$$\|x_1\| = \sqrt{1^2 + 1.1665^2} = \sqrt{2.3607} \approx 1.5366$$

Therefore:

$$x_1 = \frac{1}{1.5366}\begin{bmatrix}1\\1.1665\end{bmatrix} = \begin{bmatrix}0.6507\\0.7592\end{bmatrix}.$$

Now if we use the same method for x_2, we get

$$x_2 = \begin{bmatrix}0.7592\\-0.6507\end{bmatrix}.$$

In conclusion, $\lambda_1 \approx 3.333,\ \lambda_2 \approx -0.715,\ x_1 = \begin{bmatrix}0.6507\\0.7592\end{bmatrix}$ and $x_2 = \begin{bmatrix}0.7592\\-0.6507\end{bmatrix}$.

(b) Verification of normality:

$$\|x_1\|^2 = 0.6507^2 + 0.7592^2 = 0.4234 + 0.5764 = 1.0000$$
$$\|x_2\|^2 = 0.7592^2 + (-0.6507)^2 = 0.5764 + 0.4234 = 1.0000$$

Verification of orthogonality:

$$x_1 \cdot x_2 = 0.6507 \cdot 0.7592 + 0.7592 \cdot (-0.6507) = 0.4940 - 0.4940 = 0$$

Since the vectors are normalized and orthogonal, they are orthonormal.

(c) Let's compute $\lambda_1 x_1 x_1^T$:

$$3.333\begin{bmatrix}0.6507\\0.7592\end{bmatrix}\begin{bmatrix}0.6507 & 0.7592\end{bmatrix} = \begin{bmatrix}1.4112 & 1.6465\\1.6465 & 1.9213\end{bmatrix}$$

Let's compute $\lambda_2 x_2 x_2^T$:

$$(-0.715)\begin{bmatrix}0.7592\\-0.6507\end{bmatrix}\begin{bmatrix}0.7592 & -0.6507\end{bmatrix} = \begin{bmatrix}-0.4121 & 0.3532\\0.3532 & -0.3027\end{bmatrix}$$

Sum the results:

$$\begin{bmatrix}1.4112 & 1.6465\\1.6465 & 1.9213\end{bmatrix} + \begin{bmatrix}-0.4121 & 0.3532\\0.3532 & -0.3027\end{bmatrix} = \begin{bmatrix}0.9991 & 1.9997\\1.9996 & 1.6186\end{bmatrix}$$

Since

$$\begin{bmatrix}0.9991 & 1.9997\\1.9996 & 1.6186\end{bmatrix} \approx A$$

with an error of about 10^{-5}, which is caused by rounding, we conclude that $A = \sum_{i=1}^{2} \lambda_i x_i x_i^T$.

3. *Topic: algebra, probabilities, random variables*

Let Ω be a finite sample space, i.e., the set of all possible outcomes. Let $P(\omega) \geq 0$ be the probability of an outcome $\omega \in \Omega$. The probabilities are non-negative, and they sum up to unity, i.e., $\sum_{\omega\in\Omega} P(\omega) = 1$. Let X be a real-valued random variable, i.e., a function $X\colon \Omega \to \mathbb{R}$ which associates a real number $X(\omega)$ with each of the (random) outcomes $\omega \in \Omega$.

The expectation of X is defined by $E[X] = \sum_{\omega \in \Omega} P(\omega)X(\omega)$. The variance of X is defined by $\mathrm{Var}[X] = E\left[(X-\mu)^2\right]$, where $\mu = E[X]$.

Task a

Using the definitions above, prove that E is a linear operator.

Task b

Using the definitions above, prove that the variance can also be written as $\mathrm{Var}[X] = E[X^2] - E[X]^2$.

Hints

An operator L is said to be linear if for every pair of functions f and g and scalar $t \in \mathbb{R}$, (i) $L[f+g] = L[f] + L[g]$ and (ii) $L[tf] = tL[f]$. The proof in task b is short if you use linearity.

Solution:

(a) Let $X \colon \Omega \to \mathbb{R}$ and $Y \colon \Omega \to \mathbb{R}$ be real-valued random variables. Since

$$\begin{aligned} E[X+Y] &= \sum_{\omega \in \Omega} P(\omega)(X+Y)(\omega) \\ &= \sum_{\omega \in \Omega} P(\omega)(X(\omega)+Y(\omega)) \\ &= \sum_{\omega \in \Omega} P(\omega)X(\omega) + P(\omega)Y(\omega) \\ &= \sum_{\omega \in \Omega} P(\omega)X(\omega) + \sum_{\omega \in \Omega} P(\omega)Y(\omega) = E[X] + E[Y], \end{aligned}$$

the operator E satisfies additivity.
Let $c \in \mathbb{R}$ be a scalar. Since

$$\begin{aligned} E[cX] &= \sum_{\omega \in \Omega} P(\omega)(cX)(\omega) \\ &= \sum_{\omega \in \Omega} P(\omega)c \cdot X(\omega) \\ &= \sum_{\omega \in \Omega} c \cdot P(\omega) \cdot X(\omega) \\ &= c \cdot \sum_{\omega \in \Omega} P(\omega)X(\omega) = c \cdot E[X], \end{aligned}$$

the operator E satisfies homogeniety.
Since the opreator E satisfies additivity (i) and homogeniety (ii), E is a linear operator.

(b) Since E is a linear operator it satisfies (i) and (ii).
Furthermore, for scalar $c \in \mathbb{R}$, $E[c] = \sum_{\omega \in \Omega} P(\omega) \cdot c = c \cdot \sum_{\omega \in \Omega} P(\omega) = c \cdot 1 = c$.

Now,

$$
\begin{aligned}
\mathrm{Var}[X] &= E\left[(X-\mu)^2\right] \\
&= E\left[X^2 - 2X\mu + \mu^2\right] \\
&= E\left[X^2\right] - E\left[2X\mu\right] + E\left[\mu^2\right] && \text{(i)} \\
&= E\left[X^2\right] - 2\mu E\left[X\right] + \mu^2 E\left[1\right] && \text{(ii)} \\
&= E\left[X^2\right] - 2E[X]E\left[X\right] + E[X]^2 E\left[1\right] && \mu = E[X] \\
&= E\left[X^2\right] - 2E[X]^2 + E[X]^2 \cdot 1 && E[c] = c \\
&= E\left[X^2\right] - E[X]^2.
\end{aligned}
$$

Therefore, $\mathrm{Var}[X] = E[X^2] - E[X]^2$.

4. *Topic: conditional probabilities, Bayes rule*

The conditional probability ("X given Y") is defined by $P(X \mid Y) = P(X \wedge Y)/P(Y)$, where $P(\square)$ is the probability that $\square$ is true and X and Y are Boolean random variables that can have values of true or false, respectively. The marginal probability $P(Y)$ can also be written as $P(Y) = P(X \wedge Y) + P(\neg X \wedge Y)$, where $\neg X$ denotes logical negation.

Task a

Derive Bayes' rule $P(X \mid Y) = \frac{P(Y|X)P(X)}{P(Y)}$ by using the definition of conditional probability.

Task b

A medical test for detection of pollen allergy behaves as follows (Nevis et al. 2016):

- (i) for persons that don't have the allergy, the test gives a (false) positive in 23% of the cases, and
- (ii) for persons with the allergy, the test gives a (false) negative in 15% of the cases.

According to statistics, 20% of the population in Finland suffers from pollen allergy.

Define suitable Boolean random variables, write down the equation (in terms of the three percentages mentioned above), and compute the value for the probability that a person is allergic to pollen if we have chosen the person to test at random and the test result is positive.

Hints

You can solve task b by using Bayes' rule, the definition of conditional probability, and the expression for marginal probability mentioned above.

Solution:

(a) From basic probability, we know that $P(X \wedge Y) = P(Y \wedge X)$.

Furthermore, from conditional probability we know that $P(X \mid Y) = \frac{P(X \wedge Y)}{P(Y)}$ and $P(Y \mid X) = \frac{P(Y \wedge X)}{P(X)}$.

Now since

$$\begin{aligned}\frac{P(Y \mid X)P(X)}{P(Y)} &= \frac{\frac{P(Y\wedge X)}{P(X)} \cdot P(X)}{P(Y)} \\ &= \frac{P(Y \wedge X)}{P(Y)} = \frac{P(X \wedge Y)}{P(Y)} = P(X \mid Y),\end{aligned}$$

the Bayes' rule $P(X \mid Y) = \frac{P(Y|X)P(X)}{P(Y)}$ holds.

(b) Let $X =$"person has pollen allergy" and $Y =$"test result is positive".
We know that

- $P(Y \mid \neg X) = 0.23$
- $P(\neg Y \mid X) = 0.15 \implies P(Y \mid X) = 1 - 0.15 = 0.85$
- $P(X) = 0.20 \implies P(\neg X) = 0.80$

We want $P(X \mid Y) =$ "person has pollen allergy and the test is positive".
According to marginal probability and conditional probability,

$$\begin{aligned}P(Y) &= P(X \wedge Y) + P(\neg X \wedge Y) \\ &= P(Y \wedge X) + P(Y \wedge \neg X) \\ &= P(Y \mid X)P(X) + P(Y \mid \neg X)P(\neg X) \\ &= 0.85 \cdot 0.20 + 0.23 \cdot 0.80 = 0.354\end{aligned}$$

Now according to Bayes' rule

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)} = \frac{0.85 \cdot 0.20}{0.354} \approx 0.480.$$

Therefore, the pobability that a person is allergic to pollen and the test result is positive is 48%

5. *Topic: optimisation*

Assume you are given constants $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$, where $i \in \{1, \ldots, n\}$, and a function $f(b) = \sum_{i=1}^{n}(bx_i - y_i)^2/2$.

Task a

By using derivatives, find the value of $b \in \mathbb{R}$ that minimises the value of $f(b)$.

Task b

What conditions must the constants x_i and y_i satisfy for the function to have a unique and finite minimum?

About the topic

Optimisation is an essential tool in machine learning. For example, we try to find model parameters that minimise a given loss in model selection. Differentiation is a common way to do optimisation.

Solution:

(a) Let's first calculate the derivative $f'(b)$:

$$
\begin{aligned}
f'(b) &= \frac{d}{db}\left(\sum_{i=1}^{n} \frac{(bx_i - y_i)^2)}{2}\right) \\
&= \frac{1}{2} \cdot \frac{d}{db}\left(\sum_{i=1}^{n} (bx_i - y_i)^2\right) \\
&= \frac{1}{2} \cdot \sum_{i=1}^{n} \frac{d}{db}\left((bx_i - y_i)^2\right) \\
&= \frac{1}{2} \cdot \sum_{i=1}^{n} \left(2(bx_i - y_i) \cdot \frac{d}{db}(bx_i - y_i)\right) \\
&= \frac{1}{2} \cdot \sum_{i=1}^{n} (2(bx_i - y_i)x_i) \\
&= \sum_{i=1}^{n} ((bx_i - y_i)x_i) = b\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i.
\end{aligned}
$$

Now let's calculate the double derivative $f''(b)$:

$$
\begin{aligned}
f''(b) &= \frac{d}{db}\left(b\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i\right) \\
&= \frac{d}{db}\left(b\sum_{i=1}^{n} x_i^2\right) - \frac{d}{db}\left(\sum_{i=1}^{n} y_i x_i\right) = \sum_{i=1}^{n} x_i^2.
\end{aligned}
$$

Since $f''(b) \geq 0$, we know that $f(b)$ is convex.
Now let's solve $f'(b) = 0$ to find the minimum of $f(b)$:

$$
f'(b) = 0 \;\Leftrightarrow\; b\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i = 0 \;\Leftrightarrow\; b = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}.
$$

If $x_i \neq 0$ for at least one i, then the value b that minimises $f(b)$ is

$$
b = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}.
$$

If $x_i = 0$ for all i, the value of $f(b)$ is constant.

(b) The function $f(b)$ is constant if $\sum_{i=1}^{n} x_i^2 = 0$, otherwise $f(b)$ is strictly convex. Therefore, for $f(b)$ to have a unique and finite minimum, the constants x_i must satisfy $\sum_{i=1}^{n} x_i^2 > 0$, or in other words, $x_i \neq 0$ for at least one i.