Data Science Camp 2021 Test Tasks

# Task 1

Develop the function that computes the n-th element of Fibonacci numbers

[1, 1, 2, 3, 5, 8, 13, …] (e.g. ***fib (n=6)*** should return ***8***) and provide the value of the 60th element.

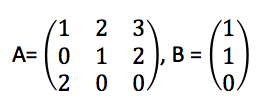
Response format:

* value of 60th element
* programming code

# Task 2

Develop program that solves the system of 3 linear equation of with 3 variables, and solve the task

𝐴𝑋=𝐵 with the help of program, where



Response format:

* Solution (replace ‘?’ with the appropriate values)
* programming code

# Task 3

The probability of outcome "H" at flipping each of the 5 coins (call them m1, m2, m3, m4, m5) with a changed center of gravity equal [0,1, 0,2, 0,4, 0,8, 0,9 ] respectively. One of the coins was chosen at random and the tests began. Determine the probability of "H" in the next flip after 8 of the actual tests: [H H H T H T H H] (here ‘T’ is oposite side of the coin). For example, before the first test, the probability of "H" is ~ 0.48 (according to the formula of full probability, taking into account the equivalence of the choice of one of the available coins). After the flip of "H" in the first test, the probability of the hypothesis that the selected coin is m1 / m2 / m3 decreased, and probabilities of hypothesis that m4 / m5, increased, and, therefore, the probability to flip "H" in the next test now equals to ~ 0.69. Similarly, after the release of "H" in another test, you need to re-compute the probability of flipping "H" in the third flip, etc.

Response format:

* list of probabilities to the nearest hundredth [0.69,?,? ,?,? ,?,? ,? ] (replace ‘?’ with the appropriate values)
* Program code, or description

# Task 4

['Data\_housing.csv'](https://drive.google.com/file/d/1J-uyJoJGjTCgp1vyi1FzBhK9mHfw6ECj/view?usp=sharing) contains records for 20,630 houses (any accommodation e.g. apartments). Attributes:

- rent: amount of monthly payment in USD

- dist: distance to center

- nrooms: number of rooms

- age: age of house

- size: are in sq.m

- noccup: number of people registered in the house

- price: price of the house in USD

**Task:**

1) The data is not completely clean. Look for anomaly values (e.g. outliers, non-number format, missed values) and try to handle them. Please describe your discoveries and selected approach.

2) Try to fill in the missed values of 'price'. Feel free to apply any approach e.g. statistical analysis or machine learning

3) Provide csv file file with data keeping all indices as original dataset. 'price' column should be number format without missing values, but instead filled with values you could predict. You may also provide values of other columns you consider more accurate than in the original dataset.

4) Provide explanation of all your findings and achievements

# Submit you work

If you have completed at least one task, or have questions, send the solutions (or a link to storage) or questions to **support@smartinsight.me.**

The test tasks have different complexity, so the total score is calculated with weights 1, 2, 3 and 5 according to the task number.

Українською

**Завдання 1.**

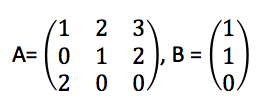
Написати програму, яка яка обчислює n-ий елемент послідовності Фібоначчі [1, 1, 2, 3, 5, 8, 13, …] (наприклад, для n = 6 програма має дати відповідь 8) та обчислити 60-ий елемент послідовності.

Формат відповіді:

* ? (замінити ‘?’ відповідним значенням)
* Код програми

**Завдання 2.**

Написати програму розв’язування систем 3 лінійних рівнянь з 3 невідомими, та вказати розв’язок X системи AX= B, де



Формат відповіді:

* Розв’язок  системи



(замінити ‘?’ відповідними значеннями)

* Код програми, котра приймає інші значення матриць A і B

**Завдання 3.**

Ймовірність випадання сторони ‘H’ для кожної з 5 монет (назвемо їх m1, m2, m3, m4, m5) зі зміщеним центром ваги рівна відповідно [0.1, 0.2, 0.4, 0.8, 0.9]. З монет навмання вибрали одну і почали випробування. Визначити ймовірність випадання ‘H в наступному випробуванні після кожного з 8 фактично проведених випробувань:

[H H H T H T H H] (тут ‘T’ протилежна сторона монети).

Наприклад, до першого випробування ймовірність випадання ‘H’ рівна ~0.48 (за формулою повної ймовірності, з урахуванням рівноможливості вибрати кожну монету з наявних). Після випадання ‘H’ в першому випробуванні, ймовірності гіпотез, що вибрана монета є m1/m2/m3 зменшились, а відповідно ймовірності гіпотез, що вибрана монета m4/m5 збільшились а, отже і змінилась ймовірність випадання ‘H’ в наступному (другому) випробуванні і стала рівною ~0.69. Аналогічно після випадання ‘H’ в другому випробування треба переоцінити ймовірність випадання ‘H’ в третьому, і т.д.

Формат відповіді:

* список ймовірностей з точністю до сотих [ 0.69, ?, ? , ?, ? ,?, ? ,? ] (замінити ‘?’ відповідними значеннями)
* Код програми, або опис/скан шляху розв’язку

**Завдання 4.**

['Data\_housing.csv'](https://drive.google.com/file/d/1J-uyJoJGjTCgp1vyi1FzBhK9mHfw6ECj/view?usp=sharing) містить записи для 20,630 будинків (будь-яке житло, наприклад, квартири), з наступними атрибутами:

- rent: сума щомісячного платежу в доларах США

- dist: відстань до центру

- nrooms: кількість кімнат

- age: вік будинку

- size: в кв.м

- noccup: кількість людей, зареєстрованих у будинку

- price: ціна будинку в доларах США

Завдання:

1) Дані можуть містити помилки (наприклад занадто великі чи відємні значення), записані у нечисловомі форматі, пропущені значення тощо. Спробуйте обробити їх, вкажіть що вдалось виявити і що зробити з такими записами.

2) Спробуйте заповнити пропущені значення "ціни". Можете застосовувати будь-який підхід, наприклад статистичний аналіз або машинне навчання

3) Надішліть файл у форматі CSV із даними, що зберігають усі індекси як і у вихідному файлі. Стовпець "price" має бути у числовому форматі без пропущених значень (заповненим значеннями, які ви можете оцінити). Ви також можете вказати значення інших стовпців, які ви вважаєте більш коректними, ніж у вхідному файлі.

4) Надішліть підсумок, що вдалось зробити

**Виконане завдання.**

Якщо ви виконали, хоч одне завдання, або у вас є питання щодо умови завдань, надсилайте виконані завдання (або лінк на їх збереження) чи питання на **support@smartinsight.me**

Запропоновані завдання мають різну складність, тому сумарний бал вираховується з ваговими коефіцієнтами 1, 2, 3 і 5 відповідно до номера завдання.