Analytic Geometry
Exercises

Exercise 1. (✎ or 💻 Norms)
Consider the vectors $(1, 2, 3, 4)$ and $(1.37, 2.74, 4.11, 5.48)$. Compute for both their Manhattan and their Euclidean norm. Compute the unit vectors (same orientation but length 1). What do you observe?

Exercise 2. (✎ or 💻 Norms)
Compute the length of vector $(4, 3, 0)$ according to both the Manhattan, the Euclidean and the Maximum norm. Explain the intuition behind the three values.

Exercise 3. (✎ or 💻 Distance Functions)
Consider the sets $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 3, 5\}$, $C = \{1, 3, 4, 5, 6\}$, $D = \{4, 5, 6\}$.

1. Compute the Jaccard-Distance between any two pairs. Which are the most similar sets?
2. Create a representation of the sets as boolean vectors. Compute the Hamming distance between any two pairs. Which are the most similar sets?

Exercise 4. (💻 Dot Product and Angles)
Find a function for the dot product in Python.

1. Verify that the vectors $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ are pairwise orthogonal
2. Compute the angle between $(1, 0, 0)$ and $(2, 0, 0)$
3. Compute the angle between $(1, 1, 1)$ and $(1, 0, 0)$

Exercise 5. (💻 Modelling Similarity)
A movie streaming service wants to cluster their users. Similar users are users who have a similar taste in movies. This is approximated by looking at the movies, users have rated.
Alice: Matrix (6 stars), Star Wars (9 stars)
Bob: Matrix (3 stars), Star Wars (4.5 stars)
Carol: Matrix (5 stars), Star Wars (5 stars)

1. Compare the Manhattan, Euclidean, Maximum and Cosine distance for pairwise comparisons of users.
2. Which metric would be the best choice for the use-case?

Exercise 6. (💻 Issues with Distance Functions)
Consider a dataset where the instances are people and the features are age and height: Alice (18y, 192cm), Bob (24y, 180cm), Carol (25y, 165cm).

1. Compute the Euclidean distance.
2. Compute the Euclidean distance when you measure the height in m instead of cm.
3. Discuss the difference of the results.
4. How can we solve this issue?