Exercise sheet 0
Setting up the software environment

Due date: As soon as possible!
Tasks: 2

In this exercise, you will learn how to set up the coding environment (at least, if you want to code in python), and how to analyze a small data set with a linear model.

1 Evacuation time versus speed

Pedestrians walk with their own desired speed. This speed is typically distributed around $1.3 m/s$, depending on the time of day, the age of the person, etc (see the book from Weidmann [2] or analyses of the speed versus the density by Seyfried et al. [1] for more details). In figure 1 you can see 100 (simulated) pedestrians walking toward a target on the right. Of course, if a pedestrian walks faster, they should reach their target faster. But how does the time to reach the target—the evacuation time for this scenario—depend on the walking speed? This is what you need to find out in this exercise.

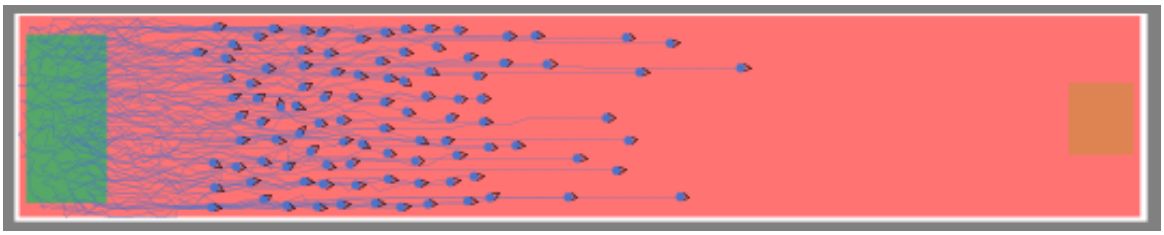

Figure 1: A simulation of pedestrians walking from the left (green field) to the right (orange box). They are visualized from the top, with the walking direction indicated by a small triangle (their "nose") and the past positions as a thin, blue line behind them.

2 Analysis with a linear model

Many relationships in the real world can be approximated well with a linear function. If we call the speed of a single pedestrian x and their individual evacuation time y, then we could ask if

$$y(x) = ax + t. \tag{1}$$

Of course, in the real world, there almost always are other effects that corrupt the measurements, for example by adding noise. But we can still test how well the simple model in (1) describes the data we measured. To do this, we have to (a) estimate the parameters a and t using parts of the data, and then (b) test if the evacuation times y in another part of the data can be described well with our new model, if we provide the speed x.

Note: the number of points per exercise is a rough estimate of how much time you should spend on each task.

Task 1/2: Setting up the environment **Points: 50/100**

Setup your implementation and simulation environment with the following:

1. Python 3.8+
2. Jupyter lab
3. PyCharm
4. Overleaf project with report template
5. Data files `speed.csv` and `endtimes.csv` from Moodle

With all of this set up, do a first exploratory data analysis of the two files. Visualize the data by plotting the evacuation times for every pedestrian against their average speed. It should look something like figure 2.

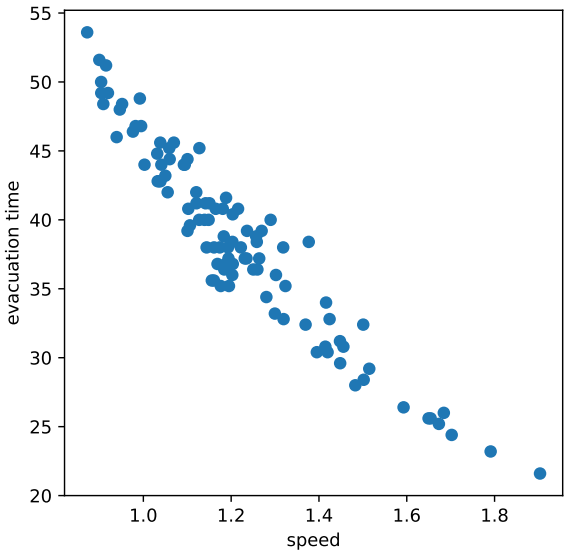

Figure 2: This is what the data should look like.

Task 2/2: Analysis with linear model **Points: 50/100**

Now, you have to test if a linear model well describes the relationship. To estimate the parameters a and t from equation (1), you should use the method of least squares. In python, you can use `numpy.linalg.lstsq` to find them. Be careful that you have to augment your data x with another column of ones, so that your model essentially is

$$y = \begin{bmatrix} x, & 1 \end{bmatrix} \begin{bmatrix} a \\ t \end{bmatrix}. \tag{2}$$

Split the data into a training set (around 70% of all the data) and test set (remaining 30%). Use the training set to determine the parameters a and t, and then test your model on the test set. Visualize your results, including the predictions of the model on train and test set.

References

[1] Armin Seyfried, Maik Boltes, Jens Kähler, Wolfram Klingsch, Andrea Portz, Tobias Rupprecht, Andreas Schadschneider, Bernhard Steffen, and Andreas Winkens. Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. In Wolfram W. F. Klingsch, Christian Rogsch, Andreas Schadschneider, and Michael Schreckenberg, editors, *Pedestrian and Evacuation Dynamics 2008*, pages 145–156. Springer Berlin Heidelberg, 2010.

[2] Ulrich Weidmann. *Transporttechnik der Fussgänger*, volume 90 of *Schriftenreihe des IVT*. Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT) ETH, Zürich, 2nd edition, 1992.