1 Visualization

Exercise 1. Let f_h denotes the frequency of class $h \in \{1, \ldots, H\}$. Verify the co-domain of the following functions:

- **Gini index** $1 - \sum_{h=1}^{H} f_h^2 \quad \in [0, (H-1)/H]$
- **Entropy index** $-\sum_{h=1}^{H} f_h \log_2 f_h \quad \in [0, \log_2 H]$
- **Miss-classification index** $1 - \max_h f_h \quad \in [0, 1/H]$

Exercise 2. Show that the eigenvalues of a symmetric matrix are always real numbers.

Exercise 3. Download the dataset `baron2016_pancreas_human_sample.csv`. Using Orange:

1. Load the file
2. Remove the variables `barcode, Cell ID, class, Batch ID, Patient, Selected`
3. Apply t-SNE
4. How many groups do you recognize?
5. Include the variable `class`
6. Notice that different groups represent the same cell class
7. Comment