Cloudflight Coding Contest
11th November 2022

Context

- Bees sometimes dance to communicate with other bees.
- Bees know several different dance moves, which can communicate different things.
- You would like to create a machine learning model that can tell you what a bee is trying to communicate with a certain sequence of dance moves.
- The bees you are observing are very good at dancing, so every bee knows exactly 9 dance moves.

- Unfortunately, bees are not so great at communicating, so they have to perform **30** dance moves in sequence to communicate a message.
- This way bees can signal 5 different messages:

0: Watch out, there is **danger.**
1: I found **food north** of the hive.
2: I found **food east** of the hive.
3: I found **food south** of the hive.
4: I found **food west** of the hive.

Input Data Formats

- There are several input files which are used for different levels.
  - **Training data:** contains labeled data to train your models.
    Every line represents one instance of training data.
    Format:
    - The dance moves of the bee represented as a string of letters (a - i).
    - The message that this sequence of moves represents as a single digit (0 – 4).
  - **Simple test data:** contains unlabeled data to test the model.
    Every line contains a string of letters representing the sequence of dance moves.
    The simple test data contains only samples from classes 0 and 1.
  - **Full test data:** has the same format as the simple test data but it contains samples from all classes.
  - **Imbalanced test data:** has the same format as the full test data but the classes are imbalanced.

LEVEL 1

Level 1: Get to know the data

Your first task is to load the **training data** and get to know the data. You should produce a histogram that shows how many samples there are of every class.

The output file should contain one line per class and every line contains a single number. Every line contains the number of samples of the corresponding class. Put simply, the first line contains the number of samples of the first class and so on.

Example:

```
100
111
122
133
144
```