Pattern Recognition
Exercise set 1

1. **[2 points]** In lectures, we went through k-nearest neighbor (kNN) classifier. How much computation you will need in classifying a test sample, assuming a data set of N training examples and dimension D?

2. **[8 points] (kNN implementation)**

   - Split the attached i-vector data into training and testing sets by using first 80 utterances (sentences) from each speaker for the training and the last 20 utterances for the testing. [about 1 point]
   - Implement kNN classifier. [about 3 points]
   - If two or more classes are equally common in the neighborhood of a test sample, reduce the value of k for this specific test case until the tie gets resolved. [about 1 point]
   - Test kNN classifier for speaker identification and gender detection tasks. For both tasks, report the classification accuracy for some value of k by comparing predicted labels to the correct ones given in the data set. [about 2 points]
   - For both tasks, plot classification accuracy as a function of k ($k = 1, \ldots, 50$). [about 1 point]

Submit your answers to Moodle by November 16, Thu, 23:55.
Late submissions:
Before November 17, Fri, 7:00: -10% of points
Before November 18, Sat, 12:00: -30% of points
After November 18, Sat, 12:00: Not accepted without a good reason that has to be given well before the deadline.

The submission should be an archive (zip, tar, etc.) that contains following files:

- `answers.pdf`: Contains answers to questions in pdf-format. If you wish, you may include scanned (readable) handwritten answers in your answer file (to avoid math typesetting). In case you give some answers in program code comments, mention it in the `answers.pdf`. Include your full name.
- `main.m` / `main.py`: A script that outputs the answers for all programming tasks. Write your full name in the first line as a comment. If using Python, please use Python3 + NumPy + matplotlib.
- Possibly other code files that your main script calls.
- Include the required data set files (ivectors.txt) in the archive so that the main script is runnable right after unpacking the archive.

Use the following naming convention for the archive file:
`<first name>_<last name>_ex<exercise set number>`
For example:
`ville_vestman_ex1.zip`