Exercise 1

In this exercise, you will implement a simple k-Nearest Neighbour classifier!

(1P) Subtask 1.1
To measure distances for a kNN classifier, it sometimes is useful to try other norms than the euclidian norm. Therefore implement a function \( pnorm(x, p) \) with \( x \in \mathbb{R}^n \) and \( p \geq 1 \):

\[
\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}
\]  

(1P) Subtask 1.2
For visualisation purposes, write a function \( cplot(T, M) \), which plots the given datasets. The matrices have the following form:

- The first column is \( \in \mathbb{N}^n \) (\( n \) being the number of samples) and contains labels
- The second and third columns represent the samples’ \( x \) and \( y \) values and are \( \in [0, 12]^n \)

Use can use Matlab’s \( \text{scatter}(...) \) and \( \text{axis}(...) \) functions. Your implementation is to plot the given datasets using 1 color per label. Make sure you plot the matrix \( M \) before \( T \), as \( M \) will be the dataset and \( T \) the training set.

(4P) Subtask 1.3
Now implement a function:

```matlab
function V = classify(T, M, k, p)
```

which classifies the gives samples \( M \in \mathbb{R}^{n \times f} \) with \( n \) being the number of samples, \( f \) the length of the feature vectors. \( V \in \mathbb{N}^n \) is a Vector containing the labeling your classifier produced. \( k \) is the number of neighbours to check for and \( p \) the parameter for the norm function. \( T \) is the already labeled training set and its structure is as described in 1.2.

(2P) Subtask 1.4
Read in the file \( train1.dat \) (\( importdata(...) \)). Now generate a matrix \( M \in [0, 12]^{2500 \times 2} \) and classify it using your classifier. Use the euclidian distance and generate 4 plots (using \( cplot(...) \)) in a single window for \( k = 1, 3, 5, 9 \).

What can you observe especially for \( k = 9 \)?

Questions: alikanso91@hotmail.com and dhhjx880713@163.com
(2P) Subtask 1.5

Read in the files `wine.train` and `wine.test`. The test set `wine.test` is labeled correctly already such that the number of errors can be measured. Remove the labels, classify the data and determine the classification error. Try to find good values for $k$ and $p$ experimentally.

Questions: alikanso91@hotmail.com and dhhjx880713@163.com