1. A binary image A and a structuring element h are shown below.
    a. Compute the erosion of h over A. The resulting image should have the same size as the original image. To do that, assume any pixel beyond the border of the image has value 0.
    b. Explain which are the effects of erosion in the mask of an undetermined object. You can base your explanation on the resulting image from $question\ a$.

$\boldsymbol{A}$

0	0	0	0	0	0	0	0	0
0	1	0	0	1	1	1	1	0
0	1	1	0	1	1	1	1	0
0	0	0	0	1	1	1	1	0
0	0	0	0	1	1	1	1	0
0	0	0	0	1	1	1	1	0
0	0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0

$\boldsymbol{h}$

1	1	1
1	1	1
1	1	1

2. An MxN data matrix is presented below, where M denotes the number of samples and N denotes the number of descriptors. Threre are also two centroids. They come from running the K-means algorithm just after a specific iteration. Starting at this point, compute the next iteration of the K-means algorithm. To do that:
    a. Do the cluster assignment of the examples to the centroids you have received.
    b. Compute the new centroids from the cluster assignment of the previous section.

$$X = \begin{bmatrix} 3 & 1 \\ 2 & 0 \\ 1 & 2 \\ 3 & 4 \\ -2 & 1 \\ -1 & 3 \\ -3 & 2 \\ 0 & 5 \end{bmatrix};\ c_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix};\ c_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

3. Consider a classification problem where the considered classes are:
    i. Apples of different colors
    ii. Bananas

    Discuss if you would use the following descriptors for the task:

    a. Mean luminance
    b. Eccentricity
    c. Perimeter of the object / Perimeter of its convex hull ratio
    d. Mean of a and b components from Lab colorspace
    e. Variance of luminance (texture)

    Assume the segmentations are always correct. Justify your answers.

4. The following data matrices are presented:

$$X_{train} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \\ -2 & -1 \\ -1 & -2 \\ 0 & 1 \\ 3 & -1 \\ 2 & 1 \\ 1 & 2 \end{bmatrix}, y_{train} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, X_{test} = \begin{bmatrix} 0 & 3 \\ -1 & 0 \\ -2 & -3 \\ 2 & -1 \end{bmatrix}, y_{test} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

It was decided to use the **Nearest Neighbor** classifier to get the predictions for the X_{test} data matrix. The chosen distance measure was the euclidean distance:

$$d(p,q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$$

Questions:

a. Compute the predictions vector for both data matrices. Remember you can draw a coordinate axis and plot the dots to evaluate the distances.
b. Compute the *accuracy* for both subsets. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

5. A set of statements is presented. Verify if each of them is true or false. Justify your answer.
   a. The most reliable evaluation measure of the performance of a classifier should be obtained from the training set.
   b. The test set can contain a few data vectors previously seen in training.
   c. We say a classifier has overfitted the data when there is a considerable difference between training and test performance.
   d. The area under the ROC curve is bounded between 0 and 1.
   e. Assume a binary classification problem. The descriptor vectors have dimension 3 and the classifier is Gaussian. To train the system, the parameters that define the class-Gaussians are to be estimated. Counting means and covariances, 18 parameters will be estimated.