**1.** The table below shows a data set **Z** with 10 objects described by 4-dimensional feature vector.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Object | *x*1 | *x*2 | *x*3 | *x*4 | Label |
| 1 | 5 | 2 | 1 | 4 | 3 |
| 2 | 6 | 1 | 4 | 0 | 2 |
| 3 | -5 | 1 | -4 | 7 | 2 |
| 4 | 0 | 1 | 3 | 2 | 2 |
| 5 | 6 | 2 | 1 | -5 | 1 |
| 6 | 4 | 2 | 1 | 7 | 1 |
| 7 | -6 | 3 | 3 | 1 | 2 |
| 8 | 0 | 2 | 1 | 8 | 3 |
| 9 | -4 | 11 | 5 | 2 | 2 |
| 10 | 4 | 7 | 2 | 2 | 3 |

**(a)** What are *C* and *N*?

**(b)** Consider a classifier *G*, which consists of the following set of discriminant functions:

*g*1(**x**) = *x*1 + *x*2 *g*2(**x**) = *x*2 + 2*x*3 *g*3(**x**) = 3*x*4 - *x*3

Run *G* on **Z** and give the (guessed) labels of all the objects. For instance, the discriminant functions for object 1 are:

g1(x) = x1 + x2 = 5 + 2 = 7

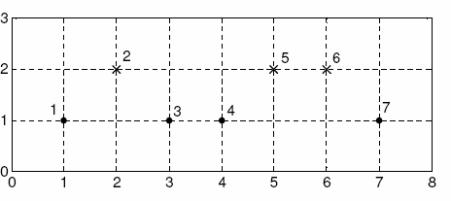
g2(x) = x2 + 2x3 = 2 + 2·1 = 4 g3(x) = 3x4 - x3 = 3·4 - 1 = 11

As max(g1, g2, g3) = g3, then the class label predicted for object 1 is 3.

**(c)** Give the confusion matrix of *G* and calculate the error of the classifier *G*.

**(d)** Estimate the prior probabilities for the classes from **Z**.

**2.** Consider the classification problem shown in the figure below. The data is depicted as points and the (true) class labels are shown by different markers: class 1 with dots and class 2 with crosses. The object numbers are given next to the markers. Your classifier operates in the following way: for a point **x** that you want to classify, find the nearest point in **Z**. Assign **x** to the class of its nearest neighbour (this is called *the nearest neighbour classifier*). In the case of a tie, resolve in favour of the class with the lower number (class 1 here).



**(a)** Find the leave-one-out error of the nearest neighbour classifier.

**(b)** Consider the following experiment. Stage1: Objects 1,2,3 and 4 are used for “training” (they form the data set from which we select the nearest neighbour) and objects 5,6 and 7 are used for testing. Stage 2: Objects 1,3,5,7 are used for training and objects 2,4,6 are used for testing. Stage 3: objects 3,4,5 and 6 are used for training and objects 1,2, and 7 are used for testing. The errors on the three testing sets are averaged to obtain an estimate of the error of the classifier. Find this estimate.

**(c)** Which training-testing protocol have we used in **(b)**?