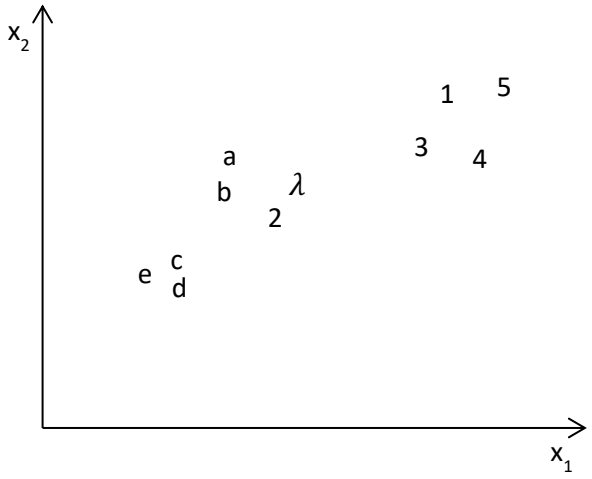
*k*NN-Classification – Exercise 1
Solutions

The diagram below shows 2-dim feature vectors of the two classes “letter” and “digit” (displayed as letters a – e and digits 1 – 5, respectively). We will apply different forms of the nearest neighbor classifier for differentiating these classes, using Euclidean distance as metric. Which class is assigned to the unknown feature vector λ in the following cases?

a) Standard 1-NN classifier
Solution: Digit

b) 3-NN classifier
Solution: Letter

c) Give criteria for rejecting (i.e. refusing to classify) feature vectors using a 3-NN classifier. Which class is assigned to the unknown feature vector x now?



Possible Solutions:
- Criterium: all three nearest neighbors have to belong to the same class. Now, x is rejected.
- Criterium: the closest neighbor must not be different from the other two. Now, x is rejected.
- Criterium: distance to the closest neighbor is larger than a threshold. Decision for x depends on the threshold now.

d) A disadvantage of the *k*NN classifier is that the complete training sample must be stored in memory and searched during classification. Propose a method to avoid this. Are there any disadvantages to your method?
Possible Solutions:
- Save only borders between classes
- Merge very close feature vectors into one
- Save only the center of each class
- Save only the border of a group of features belonging to the same class
- …
- Disadvantages: non-linear separability of classes greatly reduced in most of these approaches.