Supervised Learning
WS 22/23

Applied Artificial Intelligence
Prof. Johannes Jurgovsky

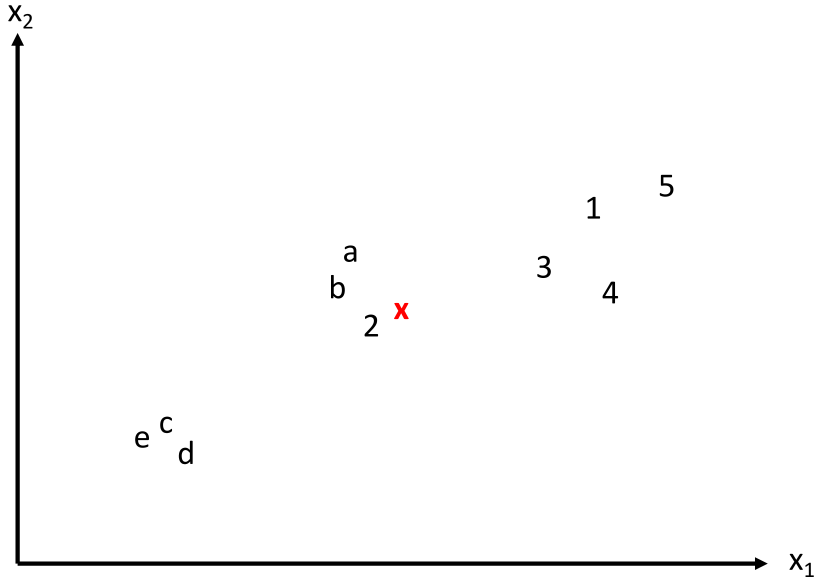
Exercise Sheet

Nearest Neighbor Learning

Learning Goals

- Nearest Neighbor Principle
- Nearest Neighbor Algorithm

1. ●○○ The diagram below shows 2D 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 x in the following cases?



   (a) Standard 1-NN classifier?

   (b) 3-NN classifier?

   (c) Suppose you want to modify kNN such that the algorithm abstains from a decision in case it is uncertain about the class. Give criteria for rejecting feature vectors using a 3-NN classifier.

   (d) A disadvantage of the kNN 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?

2. ●●○ In the attached jupyter notebook *kNN-skeleton*, you find a skeleton-implementation of the nearest neighbor classifier (Task 1). Some aspects of the implementation are missing. Read through the notebook and complete the missing bits of the algorithm:

   (a) The function `_classify(self, query_point)` predicts the class label of a single `query_point` (feature vector), taking into account the query point’s direct neighbor. Fill in the missing code.

3. ●●○ In the attached jupyter notebook *kNN-skeleton*, you find a skeleton-implementation of the k-nearest neighbor classifier (Task 2). Some aspects of the implementation are missing. Read through the notebook and complete the missing bits of the algorithm:

   (a) The function `_classify(self, query_point, k)` predicts the class label of a single `query_point` (feature vector), taking into account the query point's k nearest neighbors.

   (b) The function `predict(self, X)` receives a matrix of data points (each row is a feature vector) and returns a list of predicted class labels (your implementation should call `_classify(self, query_point, k)` repeatedly, passing each row from `X` as `query_point`.