Aufgab 9-1 Ensemble Multi-Class-Classification

We have previously considered the ensemble strategies one-versus-rest, all-pairs, and ECOC. These have allowed us to reduce multi-class classification problems to multiple two-class classification problems. For one-versus-rest und all-pairs the application/test step was a simple majority voting, ECOC required a more sophisticated decision rule.

A further approach is given by the DDAG-strategy: Individual all-pairs-classifiers form a directed, acyclic graph (DDAG=Decision Directed Acyclic Graph) to facilitate the classification result. See the following figure:

(a) What advantages and disadvantages does this strategy have compared to voting using pairwise classifiers?

(b) For each base-classifier, assume a complexity given by a function $t : \mathbb{N} \to \mathbb{R}^+_0$, which is dependent on the number of training samples. How do different strategies perform regarding the time requirements in the training phase for $n$ classes and $m$ samples in each class? How do they perform in the application phase, assuming constant time for a prediction of an single base-classifier?
Aufgabe 9-2  Complementarity of Classifiers

Let \( f_1 \) and \( f_2 \) be two binary classifiers, which predict a class \( c \in \{0, 1\} \) given a representation of an object from a dataset \( D \). Decide whether they should be combined if:

(a) \( f_1(x) = f_2(x) \) for all \( x \in D \)

(b) \( f_1(x) = 1 - f_2(x) \) for all \( x \in D \)

Aufgabe 9-3  Measure of Dependency

Let \( h \) be a measure of the dependency between two kernel matrices \( K \) and \( K' \). That is \( h(K, K') \) is large if the associated kernels \( k \) and \( k' \) consider the same objects as similar (and dissimilar). If they consider the similarity between the same objects differently, \( h(K, K') \) is low.

Further, let \( D \) be a data set with class labels and \( r \) representations per object. We calculate a kernel matrix \( K_i \) for each representation and a kernel matrix \( L \) on the class labels. Describe how \( h \) can be used to determine a linear combination of the \( K_i \), which reflects the similarity of the class labels as well as possible.

Aufgabe 9-4  Multi-Represented Classification

Given a dataset with multiple representations of each data object. We want to determine their class association using multiple representations.

- Which phase of the classification process is best suited to integrating different representations.
- How can multiple representations be incorporated into the training phase?
- How can multiple representations be incorporated into the prediction phase?
- Is it necessary to normalize the objects before each phase?

Aufgabe 9-5  Multiview Clustering

Given a dataset \( X \) such that each point is represented by two two-dimensional vectors:

\[
A = (0, 1); (3, 0) \\
B = (-1, -1); (2, 0) \\
C = (0, 0); (3, 1) \\
D = (0, -3); (-2, 2) \\
E = (2, 1); (-2, -3)
\]

We want to realize a multiview clustering using DBSCAN.

(a) How does multiview clustering differ from ordinary clustering? Which particular difficulties does it face?

(b) Let \( \text{minPts} = 3 \). Which values for \( \varepsilon_1, \varepsilon_2 \) make objects \( C \) and \( D \) core objects using the

- union method?
- intersection method?