
11 Conclusion

The main goal of this dissertation was the development of new learning algorithms for crisp and soft k-nearest-neighbour classifiers in the light of current advances in machine learning like large margin classifiers or ensemble learning. The thesis also explores global training algorithms for pattern recognition based on a feature extractor that uses several lineal combinations of input variables.

We have obtained the following results, listed according to their relative importance:

1. Development of learning algorithms for crisp and soft k-NN classifiers that have a large margin distribution of correct classified training patterns.
2. Elaboration of a novel *oriented principal component analysis* (OPCA) addressed for finding optimal feature extraction in classification problems based on a global gradient training of the feature extractor and the classifier.
3. Development of *local stabilization techniques for ensembles of NN classifiers*. The two new methods addressed for NN ensembles use local stabilisation techniques to derive a single predictor that have the same complexity than each single member of the ensemble.
4. Extension and generalization of Kohonen's LVQ algorithms that include:
 - 4.1. A simple generalization of LVQ1 that improves the probability of reaching a minimum of the training error. We derive the on-line version and a batch version based on Newton optimisation.
 - 4.2. A family of batch LVQ algorithms that implements the supervised clustering of Kohonen's LVQ algorithms in a more principled way
 - 4.3. A dynamic LVQ algorithm to address the problem of determining proper initial values for these algorithms and the optimal number of prototypes.
5. Study of the finite-sample convergence of the on-line LVQ1 and K-means algorithms. This novel analysis computes the real equilibrium points of the on-line learning equations and

derives the conditions to ensure good convergence and small optimization error using tools of the dynamic system and optimization theories.

We started with the study of some existing approaches based on clustering like the K-means and Kohonen's LVQ1. We analysed the finite-sample (or real) convergence of these algorithms (chapters 3 and 4).

Then, we extended LVQ1 in the light of the results of the finite-sample convergence. We derived a new version of the LVQ1 algorithm using the Newton optimization to avoid optimization errors and to ensure faster convergence. Besides, we derived a generalization of LVQ1 (GLVQ1) which regulates the degree of punishment of the learning algorithm when a training sample is assigned to wrong class (chapter 4). GLVQ1 includes the LVQ1 and k-means as a special cases.

GLVQ1 is a very simple extension that can improve the classification accuracy. However, it has the drawback that its regularising parameter must be empirically determined. Then, a family of algorithms was obtained using the underlying idea that Kohonen's LVQ algorithms use in a more principled way. We implemented a learning algorithm based on Newton optimization (BLVQ) that performs a clustering process over a modified probability density function which is zero at Bayes borders (chapter 5).

The problem of determining proper initial values for these LVQ algorithms and the optimal number of prototypes was then studied and a dynamic LVQ algorithm was proposed (chapter 6).

The issue of stabilising an ensemble of NN classifiers trained with these algorithms was also addressed. Two methods based on the local nature of NN classifiers were proposed to achieve a predictor with a similar number of parameters than the single members of the ensemble (chapter 7).

We then derived new learning algorithms for crisp and soft k-NN classification (like the Learn1NN algorithm) that minimises the training error and also achieves a large margin distribution on the training samples. The resulting classifiers were accordingly called *large margin nearest neighbour classifiers* (chapters 8 and 9). Interesting relationships between large margin NN classifiers and SVM were pointed out in our work.

Finally, the thesis proposed a novel method (called *oriented principal component analysis*, OPCA) to perform a *global gradient-based training* of a feature extractor that uses several lineal combinations of input variables and any classifier that allow a back-propagation of an error signal through its architecture (chapter 10). The idea of the method is to perform a series of

training sessions until it finds an optimal projection of the input variables in the feature space that allows a better separation of classes.

More than 150000 lines of C code were developed to implement the learning algorithms and other functions needed to test and manipulate different statistics of the NN classifiers. The LVQ_PAK library was used as the core library of our own libraries.

The work described here could be extended in different ways. Among them are:

1. *Use of more sophisticated VQ schemes and learning algorithms to derive new classifiers based on data-dependent partitioning.* A classifier based on data-dependent partitioning divides the input space in cells according to a rule. This thesis is mainly devoted to nearest-neighbour cells, one of the simplest forms of partition. Advanced VQ based on modular and tree structures could be employed to derive new forms of classification. Besides, more complex clustering learning algorithms like the EM (expectation-maximization) algorithm or fuzzy clustering algorithms (e.g. soft k-means) could be employed in nearest-neighbour partitioning and also in the advanced VQ schemes.
2. *Use of advanced clustering algorithms in local averaging.* Local averaging makes use of K-means to compute cluster centroid. However, there exist more advanced clustering algorithms for this task like soft K-means or EM.
3. *Study of the theoretical properties of large margin nearest-neighbour classifiers.* The large margin NN classifiers derived in this thesis minimise the number of training errors as the result of maximising the margin of correct classifications. Consequently, capacity of these classifiers does not depend on the *VC dimension* but a scale sensitive version called *fat-shattering*. Generalization error bounds could be then derived if the fat-shattering was computed.
4. *Use of Mercer's kernels to derive NN classifiers that project input data to high-dimensional feature spaces.* As we have shown, large margin classifiers and support vector (SVM) machines are related. SVM transform the original input space into a high-dimensional space where it places a single optimal margin hyperplane (OH). However, by the use of *kernels*, all necessary computations are only performed in the input space. On the other hand, Learn1NN-based NN classifiers places a series of OH's in the input space. It is possible that large margin NN classifiers also make use of *kernels*. Since, a classification problem has more chances of being linearly separable when a high-dimensional transformation is performed, kernel-based NN classifiers could employ then a smaller number of OH's to solve the classification problem with better generalization error.

5. *Derivation of local OPCA.* OPCA uses a global linear transformation to project input data on a feature space. However, performing OPCA in different regions of the input space can enhance the classification performance.