



## Novel classification using adaptive metric k-NN

Chen, Chen; Chernoff, Konstantin; Karemore, Gopal; Lo, Pechin Chien Pau; Nielsen, Mads; Lauze, Francois Bernard

*Publication date:*  
2011

*Document version*  
Peer reviewed version

*Document license:*  
[Unspecified](#)

*Citation for published version (APA):*  
Chen, C., Chernoff, K., Karemore, G., Lo, P. C. P., Nielsen, M., & Lauze, F. B. (2011). *Novel classification using adaptive metric k-NN*. Abstract from 1st IEEE Workshop on Kernels and Distances for Computer Vision in conjunction with ICCV 2011, Barcelona, Spain.

# Novel classification using adaptive metric $k$ -NN

C. Chen, K. Chernoff, G. Karemore, P. Lo, M. Nielsen, and F. Lauze  
Department of Computer Science, University of Copenhagen, Denmark

## 1. INTRODUCTION

The  $k$ -nearest-neighbors ( $k$ -NN) algorithm has been widely used as one of the simplest and model-free supervised classifier. It requires a distance function in feature space in order to identify the  $k$  nearest neighbors, and its performance is expected to be highly influenced by the specific choice of metric. The metric used is often chosen as the standard Euclidean metric, reflecting the absence of any specific prior on the data. However, especially for high-dimensional data, this choice can contribute to the deterioration of the classifier, as it assumes that the data points are distributed in a hyperspherical manner in feature space and independent of the class labels, which might not always be optimal in terms of class-label discrimination.

In recent years, it has been shown that using an appropriate metric can yield significant improvements when compared to the Euclidean metric.<sup>1</sup> In this work, we studied some special metrics, all from the simple class of generalized Euclidean ones, defined by a symmetric (semi) positive definite (SPD or SSPD) matrix  $M$  on the feature space

$$d_M(x, y)^2 = (x - y)^t M (x - y). \quad (1)$$

The metrics under investigation are three Mahalanobis distance based metrics and a metric obtained through supervised learning. The first ones reflect typically some knowledge or a priori on the distribution of unlabeled data, while the last one takes explicitly into account labeled data.

## 2. ADAPTIVE METRIC $k$ -NN

### 2.1 Brownian Image Model

A well studied family of models for natural images are the so called Fractional Brownian Image Models introduced by Mandelbrot.<sup>2</sup> According to Brownian Image Model theory, all the partial derivatives of a Brownian image at any given location  $(x, y)$  are Gaussian distributed. Covariance of the outcome of the different filters can be computed analytically:

$$E[\langle \partial_{x^{m_1} y^{n_1}} G_{s_1} * I, \partial_{x^{m_2} y^{n_2}} G_{s_2} * I \rangle] = (-1)^{\frac{m+n}{2} + n_2 + m_2} \frac{\sigma_0^2}{2\pi} \frac{n!m!}{2^{n+m} (n+m) (\frac{n}{2})! (\frac{m}{2})!} \quad (2)$$

whenever both  $m = m_1 + m_2$  and  $n = n_1 + n_2$  are even, otherwise the covariance is 0. The resulting metric defined by Eq. (1) is with matrix  $M_{BIM} = \Sigma_{BIM}^{-1}$  and we will refer to it as the “theoretical metric”.

### 2.2 Normalized metric

The normalized metric situation is a very commonly used one, where one assumes that features are statistically independent. The covariance matrix  $\Sigma_N$  reduces to the diagonal matrix of variances of individual features. These variances are computed empirically:  $(\Sigma_N)_{ii} = \text{Var}(x^i)$ ,  $i = 1, \dots, N$ , where  $x^i$  denotes the  $i$ -th component of the  $N$ -dimensional feature vector  $x$ . The resulting metric, as defined by Eq. (1), will be denoted as  $M_N = \Sigma_N^{-1}$ . We will refer it as the “normalized metric” or “standardized metric”.

### 2.3 Empirical metric

With Gaussian assumptions on data set, the covariance can be estimated via the classical empirical formula:

$$\Sigma_{EMP} = \frac{1}{U} \sum_{i=u}^U (x_u - \mu)(x_u - \mu)^T, \quad (3)$$

where  $x_u$ ,  $u = 1, \dots, U$  are sample points in the data set,  $U$  is the number of sample points, and  $\mu$  is the empirical mean of the data set. The resulting metric is referred as “empirical metric”, which reflects the structure of the unlabeled training data.

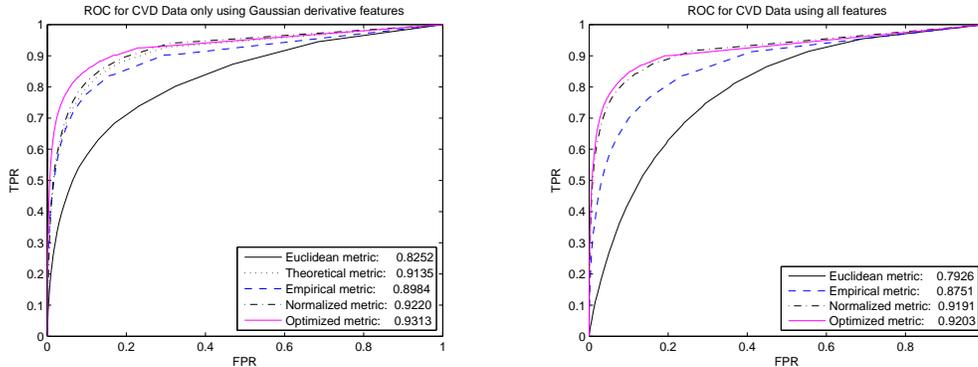


Figure 1: The ROC curves and area under ROC curves of classification results using the standard Euclidean metric and four adaptive metric (left: only using Gaussian derivative features; right: using all the features).

## 2.4 Optimized metric

As opposed to the previous discussed cases, this section introduces a metric obtained through minimization of a cost-functional corresponding to the weighted Leave-One-Group-Out (LOGO) error of  $k$ -NN.

Given a sample set  $D$ , we assume it is naturally partitioned via “group”  $p$ , and let  $D_p$  denote the subset of samples that “belong to group”  $p$  and  $\hat{D}_p$  its complement  $D \setminus D_p$ . The  $u$ -th sample for group  $p$  is denoted  $x_{(u,p)}$ , and its label is denoted by  $l_{(u,p)}$ . The weighted  $k$ -NN LOGO error as:

$$E(M; D, k) = \sum_{(u,p)} \sum_{n=1}^k w_n \delta \left( \text{NN}_n(x_{(u,p)}, \hat{D}_p), l_{(u,p)} \right), \quad (4)$$

where  $\text{NN}_n(x, S)$  is the label of  $n$ -th nearest neighbor of  $x$ , and  $w_n$  is defined by  $\sum_{j=1}^k C_U^j \xi^j (1 - \xi)^{U-j}$ , where  $\xi = \frac{n}{U}$  and  $U$  is the cardinality of the data set. The error is approximated by a differentiable smooth functional and optimized using a stochastic version of the gradient descend algorithm.

## 3. EXPERIMENT & RESULT

We applied the proposed methods on cardiovascular disease (CVD) data for detecting calcifications in the lumbar section of the aorta in X-ray images. The data set consists of 14 images, and a total of 52 features at each of 5000 points were extracted from each image: the image intensity, gradient magnitude, Hessian based features, and Gaussian derivatives at total order up to 3 at three different scales.

Figure 1 shows the classification results of the Receiver Operating Characteristic (ROC) curves and the area under the ROC curves (numbers in the bottom boxes). As observed from the figure, all the proposed methods perform better than the standard Euclidean metric. The optimized metric outperforms the rest. Among the rest three methods, which method performs the best depends on the training data.

## 4. CONCLUSION

We discussed adaptive metric  $k$ -NN for classification based on different prior knowledge: three Mahalanobis distance based metrics and a optimized metric obtained via minimizing a  $k$ -NN specific classification error. Experiment on CVD data showed that  $k$ -NN classifier benefited greatly from the proposed metrics as compared to the standard Euclidean one. The adaptive metric can be also applied to other distance based classifiers.

## REFERENCES

1. E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, “Distance Metric Learning with Applications to Clustering with Side Information,” in *Advances in Neural Information Processing Systems*, **15**, pp. 505–512, 2003.
2. B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, San Francisco, CA, 1982.