On the Reliability of kNN Classification

Maxim Tsypin and Heinrich Röder

Abstract—**We propose a formula that quantifies the reliability of kNN classification in the two class case. Let the training set consist of N_1 instances of class 1 and N_2 instances of class 2. Let the k nearest neighbors of the test instance contain k_1 instances of class 1 and k_2 instances of class 2, $k = k_1 + k_2$, $k_1 << N_1$, $k_2 << N_2$. We derive, under some additional assumptions, the estimate for the probability that the test instance belongs to class 1.**

Index Terms—**classification, kNN, confidence, probability, error**

I. Introduction

k-Nearest Neighbor (kNN) is a powerful method of nonparametric discrimination, or supervised learning [1]-[4]. Each object, or instance, to be classified, is characterized by d values $x_i, i = 1 \ldots d$ and is thus represented by a point in d-dimensional space. The distance between the two instances can be defined in different ways, the simplest of which is the usual Euclidean metric $\sqrt{\sum_i (x_i - x_i')^2}$. Given a training set (a set of instances with known class assignments) and a positive (odd) integer *k*, classification of the test object is performed as follows.

1. In the training set, find the k nearest neighbor training instances to the test object.
2. Each of these k instances belongs to one of the two classes. As k is assumed odd one class has more members in the k-neighborhood of the test instance.
3. Classify the test instance as belonging to this class.

This simple algorithm has two noticeable drawbacks. First, it does not properly take into account the number of instances of each class in the training set. Simply adding more instances of a given class to the training set would bias classification results in favor of this class. Thus the algorithm in the above simple form is only applicable when each class in the training set is represented by an equal number of instances.

Second, the algorithm provides no information on the confidence of class assignment for individual test instances. Consider, for example, the case of $k = 15$ and two classes. It is intuitively clear that the confidence of class assignment in the 15:0 situation is much higher than in the 8:7 situation. In many applications, such as those related to clinical diagnostics, it is very important to be able to characterize the confidence of each individual class assignment.

Here we address these problems by providing a probability estimate of the test instance belonging to each of the classes, based on the class labels of each of the *k* nearest neighbors from the training set. We restrict ourselves to the case of two classes. We provide two derivations, one within the kernel density estimation framework (a fixed vicinity of the test instance determines the number of neighbors), the other within the kNN framework (a fixed number of neighbors determines the size of the vicinity). Both lead to the same result for the probability estimate of the test instance belonging to each of the classes.

II. Probability Estimate Within the Kernel Approach

Consider the case of two classes, which we denote as 1 and 2. Each instance is represented by a point $\vec{x} = (x_1 \ldots x_d)$ in a metric d-dimensional space. Denote the full d-dimensional space by Ω. Class 1 is characterized by the (unknown) probability distribution $p_1(\vec{x}), \int_\Omega p_1(\vec{x})\,d\vec{x} = 1$. Class 2 is characterized by the (unknown) probability distribution $p_2(\vec{x}), \int_\Omega p_2(\vec{x})\,d\vec{x} = 1$. The training set consists of N_1 points drawn from class 1, and N_2 points drawn from class 2. Denote a vicinity of the test point (representing the test instance) by ω. (In case the Euclidean distance is used, ω is a sphere centered at the test point, but this is irrelevant for the following.) For a given realization of the training set, we observe k_1 points in ω from class 1, and k_2 points in ω from class 2.

We make the following assumptions and approximations.

1. *Existence of probability densities*. As already stated above, the training set is considered to be a sample drawn from the underlying probability distributions with densities $p_1(\vec{x})$ and $p_2(\vec{x})$.
2. *Fixed ω.* The vicinity ω of the test point is considered fixed: It can depend on the position of the test point, on the probability distributions from which the training set is drawn, as well as on N_1 and N_2, but is assumed to stay the same for each realization of the training set.
3. *Uniformity within ω.* The vicinity ω is sufficiently small, such that probability densities for the classes, $p_1(\vec{x})$ and $p_2(\vec{x})$, are approximately constant within ω.
4. *Poisson approximation.* For each class, we make an approximation that the number of training set instances of this class in ω is drawn from the Poisson distribution. This

Manuscript received July 22, 2007. M. Tsypin (phone: 970-846-6953; fax: 970-870-9045; e-mail: mtsypin@biodesix.com) and H. Röder (email: hroder@ biodesix.com) are with Biodesix, Steamboat Springs, CO 80487 USA

approximation is valid when $k_i << N_i$, $\int_\omega p_i(\vec{x})\,d\vec{x} << 1$, $i = 1, 2$.

In the Poisson approximation, k_i is drawn from the Poisson distribution with expectation value λ_i,

$$\lambda_i = N_i \int_\omega p_i(\vec{x})\,d\vec{x},\ i = 1, 2.$$

Assuming equal prior probabilities for class assignment of the test point, that is, $P(\text{class 1}) = P(\text{class 2}) = 0.5$ in the absence of any information about the neighbors, the probabilities of the test point belonging to class 1 or to class 2 are as follows:

$$\frac{P(\text{class 1})}{P(\text{class 2})} = \frac{\int_\omega p_1(\vec{x})\,d\vec{x}}{\int_\omega p_2(\vec{x})\,d\vec{x}}.$$

Thus

$$P(\text{class 1}) = \frac{\int_\omega p_1(\vec{x})\,d\vec{x}}{\int_\omega p_1(\vec{x})\,d\vec{x} + \int_\omega p_2(\vec{x})\,d\vec{x}} = \frac{\lambda_1 / N_1}{\lambda_1 / N_1 + \lambda_2 / N_2}.$$

Here we have also implicitly used the uniformity assumption (assumption 3). Now we can estimate λ_1 and λ_2 in the usual Bayesian manner. Both k_1 and k_2 are assumed to obey the Poisson distribution,

$$p(k \mid \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Denoting the prior distribution for λ by $p_0(\lambda)$, by standard Bayesian reasoning we obtain

$$p(\lambda \mid k) = \frac{p(k \mid \lambda) p_0(\lambda)}{\int d\lambda\, p(k \mid \lambda) p_0(\lambda)}.$$

Assuming from now on a flat prior distribution of λ, $p_0(\lambda) = const$, we obtain

$$p(\lambda \mid k) = p(k \mid \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Eventually,

$$P(\text{class 1}) = \int_0^\infty d\lambda_1 \int_0^\infty d\lambda_2 \frac{\lambda_1}{\lambda_1 + (N_1 / N_2)\lambda_2} p_1(\lambda_1)\, p_2(\lambda_2),$$

where

$$p_1(\lambda_1) = \frac{\lambda_1^{k_1}}{k_1!} e^{-\lambda_1},\ \ p_2(\lambda_2) = \frac{\lambda_2^{k_2}}{k_2!} e^{-\lambda_2}.$$

Computation of this integral gives

$$P(\text{class 1}) = \frac{k_1 + 1}{k_1 + k_2 + 2} \cdot {}_2F_1\left(1, k_2 + 1; k_1 + k_2 + 3; 1 - \tfrac{N_1}{N_2}\right). \quad (1)$$

This is our main result. For equal sample sizes in the training set $(N_1 = N_2)$ this simplifies to the following:

$$P(\text{class 1}) = \frac{k_1 + 1}{k_1 + k_2 + 2},\ \ \frac{P(\text{class 1})}{P(\text{class 2})} = \frac{k_1 + 1}{k_2 + 1}. \quad (2)$$

III. Probability Estimate Within the kNN Framework

The above estimate has been obtained in the fixed-ω framework. It is interesting to check whether it holds in the fixed-*k*, i.e. proper kNN framework. For each of the *k* nearest neighbors we have,

$$\tilde{p}_1 \equiv P(\text{neighbor belongs to class 1})$$
$$= \frac{N_1 \int_\omega p_1(\vec{x})\,d\vec{x}}{N_1 \int_\omega p_1(\vec{x})\,d\vec{x} + N_2 \int_\omega p_2(\vec{x})\,d\vec{x}}. \quad (3)$$

Then

$$P(k_1 \text{ of } k \text{ NN belong to class } 1 \mid \tilde{p}_1) = \binom{k}{k_1} \tilde{p}_1^{k_1} (1 - \tilde{p}_1)^{k - k_1},$$

where NN stands for “nearest neighbors”. Thus

$$P(\tilde{p}_1 \mid k_1 \text{ of } k \text{ NN belong to class 1}) =$$
$$= \frac{P(k_1 \text{ of } k \text{ NN belong to class } 1 \mid \tilde{p}_1) P_0(\tilde{p}_1)}{\int P(k_1 \text{ of } k \text{ NN belong to class } 1 \mid \tilde{p}_1) P_0(\tilde{p}_1)\, d\tilde{p}_1}$$

$$= \frac{\binom{k}{k_1} \tilde{p}_1^{k_1} (1 - \tilde{p}_1)^{k - k_1}}{\int_0^1 \binom{k}{k_1} \tilde{p}_1^{k_1} (1 - \tilde{p}_1)^{k - k_1}\, d\tilde{p}_1} = (k+1) \binom{k}{k_1} \tilde{p}_1^{k_1} (1 - \tilde{p}_1)^{k - k_1}, \quad (4)$$

where we have again assumed a flat prior distribution of $\tilde{p}_1$: $P_0(\tilde{p}_1) = const$. From (3) we obtain

$$\frac{\int_\omega p_2(\vec{x})\,d\vec{x}}{\int_\omega p_1(\vec{x})\,d\vec{x}} = \frac{N_1}{N_2}\left(\frac{1}{\tilde{p}_1} - 1\right).$$

Then for given $\tilde{p}_1$ the probability that the test point belongs to class 1 is

$$P(\text{class 1}) = \frac{1}{1 + \frac{\int_\omega p_2(\vec{x})\,d\vec{x}}{\int_\omega p_1(\vec{x})\,d\vec{x}}} = \frac{N_2 \tilde{p}_1}{N_2 \tilde{p}_1 + N_1 (1 - \tilde{p}_1)}.$$

Finally, using (4),

$$P(\text{class 1}) = \int_0^1 \frac{N_2 x}{N_2 x + N_1 (1 - x)} (k+1) \binom{k}{k_1} x^{k_1} (1 - x)^{k - k_1}\, dx$$
$$= \frac{k_1 + 1}{k_1 + k_2 + 2} \cdot {}_2F_1\left(1, k_2 + 1; k_1 + k_2 + 3; 1 - \tfrac{N_1}{N_2}\right), \quad (5)$$

which is identical to the result (1) obtained in the previous section.

IV. Conclusion

Unlike the estimates of the overall error rate of kNN classification [1]-[3] that depend on the probability distributions associated with the classes, our formula (1) provides a reliability of class assignment for each individual test instance, depending only on the (known) training set data. It also properly accounts for complications arising when the numbers of training instances in the two classes are different, i.e. $N_1 \neq N_2$.

The problem of statistical confidence of kNN classification has been also addressed in [5]-[8]. However, the “confidence level” proposed in [5]-[7] has a completely different statistical meaning and cannot be used to estimate the reliability of class assignment for each individual test instance. The same is true for *P*-values discussed in [8], p.466.

REFERENCES

[1] E. Fix and J. L. Hodges, “Discriminatory analysis. Nonparametric discrimination: consistency properties.” Report Number 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas (February 1951). Reprinted in *International Statistical Review*, 57 (1989) 238-247.

[2] E. Fix and J. L. Hodges, “Discriminatory analysis. Nonparametric discrimination: small sample performance.” Report Number 11, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas (August 1952).

[3] T. M. Cooper and P. E. Hart, “Nearest Neighbor Pattern Classification”, *IEEE Transactions on Information Theory*, IT-13 (1967) 21-27.

[4] B. W. Silverman and M. C. Jones, “E. Fix and J. L. Hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation. Commentary on Fix and Hodges (1951).” *International Statistical Review*, 57 (1989) 233-238.

[5] J. Wang, P. Neskovic and L. N. Cooper, “Partitioning a feature space using a locally defined confidence measure”, ICANN/ICONIP (2003) 200-203.

[6] J. Wang, P. Neskovic and L. N. Cooper, “An adaptive nearest neighbor algorithm for classification”, Proceedings of the 4th International Conference on Machine Learning and Cybernetics, Guangzhou (2005) 3069-3074

[7] J. Wang, P. Neskovic and L. N. Cooper, “Neighborhood size selection in the *k*-nearest-neighbor rule using statistical confidence”, *Pattern Recognition* 39 (2006) 417-423.

[8] X.-J. Ma, R. Patel, X. Wang, *et al*, “Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay”, *Arch. Pathol. Lab. Med.* 130 (2006) 465-473.