Implementation of *One*-Nearest Neighbor, *k*-Nearest Neighbor, Bayes and Parzen Window Classifiers

Majid Nasiri Manjili
Shahid Rajaee teacher training University, Tehran, Iran
majid.nasiri@srttu.edu

Abstract

In this report different classifiers include one-nearest neighbor, k-nearest neighbor, weighted k-nearest neighbor, bayes and Parzen window are implemented on different datasets (synthetic 2d gaussian, iris and satimage) related to machine learning course.

Keywords: classification, 1NN, kNN, wkNN, bayes, Parzen window

1 Introduction

Classification is task of assigning a class to unknown samples, this is a vital part of every machine learning problem. In fact better classification results more accuracy in following parts of algorithm and finally higher performance. This is the reason why the improvement of this classifiers has remained an active research topic in past years.

In his report different classifiers like one-nearest neighbor (*1*NN), *k*-nearest neighbor (*k*NN), bayes and Parzen window are implemented on different datasets (synthetic 2d gaussian distribution , iris and satimage) related to machine learning course.

In the following parts brief information about datasets is noted (2), and after developing technical details of classifiers (3) finally there will be conclusion and comparing different classifiers (4).

2 Datasets

Three different datasets have been used to evaluate classification ability of classifiers. This first one is a two class named positive class and negative class by normal distribution and two features (Figure 1).

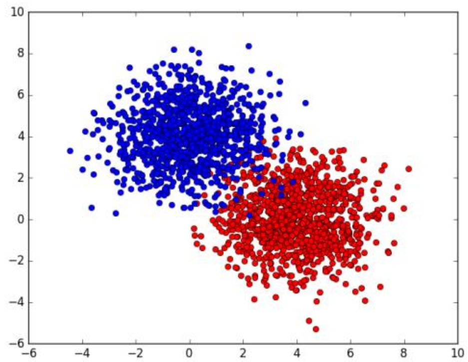

Figure 1: Normal distribution of two classes

second one is iris datasets that consists of 3 different types of irises' (Setosa, Versicolour, and Virginica) petal and sepal length, stored in a 150x4 array, The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width [1]. The third dataset is satimage, this dataset consists of the multi-spectral values of pixels in 3x3 neighborhoods in a satellite image, and the classification associated with the central pixel in each neighborhood. In this dataset there are 6 classes and 6435 samples (4435 training set and 2000 test set) with 36 attributes for every samples.

3 Classifiers

3.1 *One*-Nearest Neighbor

The closest neighbor rule distinguishes unknown samples on the basis of its closest neighbor whose class is already known. We have applied this technique on our three datasets and we achieved high accuracies **99.0%**, **95.33%** and **64.9%** on datasets synthetic-normal distribution, iris and satimage respectively. For more robust result 10 cross-validation technique have used.

3.2 *k*-Nearest Neighbor

K-nearest neighbor algorithm is computed on the basis of estimation of k that indicates how many nearest neighbors are to be considered to characterize class of a sample data point. The three following Figures 2, 3, and 4 are showing results of implementation this algorithm on datasets.

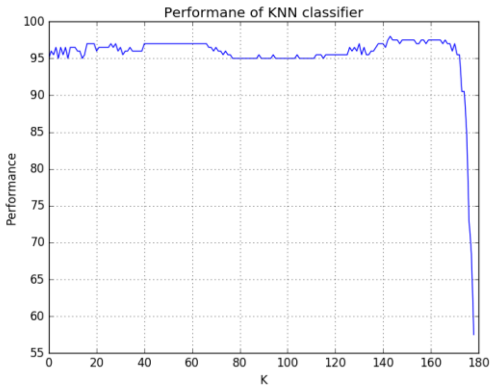

Figure 2: Performance of *k*NN classifier on synthetic normal distribution

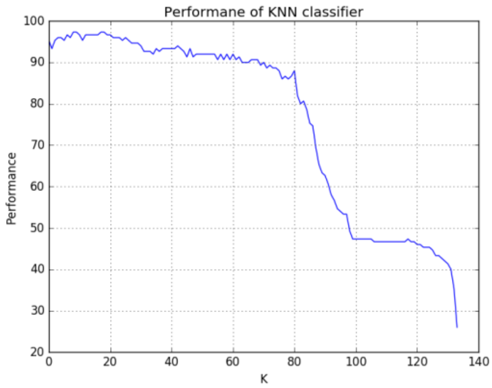

Figure 3: Performance of *k*NN classifier on iris dataset

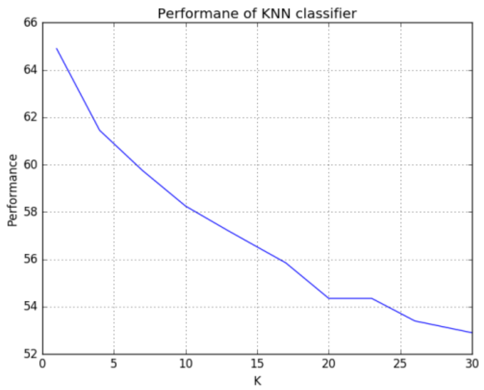

Figure 4: Performance of *k*NN classifier on satimage dataset

3.3 Weighted *k*-Nearest Neighbor

The weighted *k*-Nearest Neighbor classifier do better job in comparison with *k*-Nearest Neighbor, this version of *w*-*k*NN chooses those class that their samples are closer to test sample, when there are equal number of samples in *k* closest samples. We only have tested this classifier on iris dataset and draw their performance diagram in a figure (figure 5).

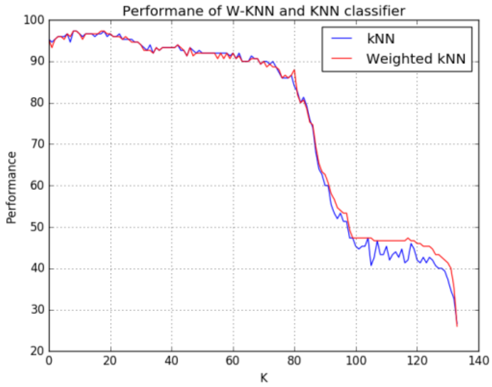

Figure 5: comparing of *k*NN and *wk*NN classifier on iris dataset

3.4 Bayes

The Bayes Classifier technique is based on Bayesian theorem and is capable of calculating the most possible output based on the input. Bayes theorem provides a way of calculating the posterior probability, $P(w_i|x)$ from $P(w_i)$, $P(x)$ and $P(x|w_i)$ using equation 1.

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)} \quad (1)$$

$P(w_i|x)$ is the posterior probability of class given attribute of class, $P(w_i)$ is called the prior probability of class, $P(x|w_i)$ is the likelihood which is the probability of predictor of given class and $P(x)$ is the prior probability of predictor of class.

In the first test, we have used two (red class & blue class) 2d normal distributions for two different classes and employed bayes classifier to determine which classes are test samples belong to, and the result are promising and something around **97.55%**. Further we have continued to classify iris and satimage datasets using Bayes classifier we reached the performance of **97.33%** and **85.18%** respectively.

3.5 Parzen window

New data in Parzen window classifier is belong to maximum number of data with same class in a space surrounded by a ball. In Parzen window classifier, h is the window width or bandwidth parameter that corresponds to the width of the kernel. The bandwidth h is typically chosen based on the number of available observations n. Figures 11, 12 and 13 shows the performance of Parzen window classifier with different h on datasets.

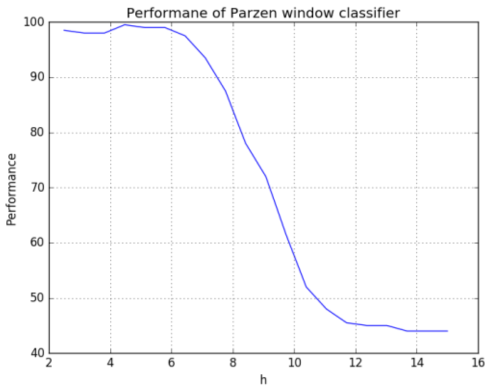

Figure 6: Performance of Parzen window classifier on synthetic normal distribution

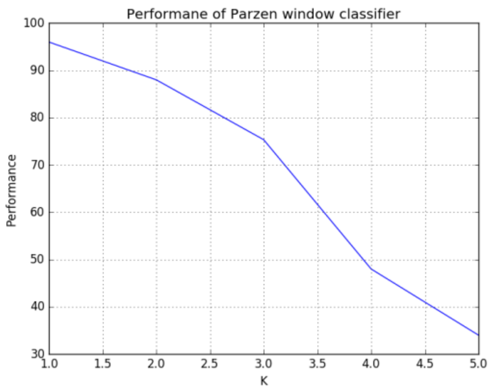

Figure 7: Performance of Parzen window classifier on iris dataset

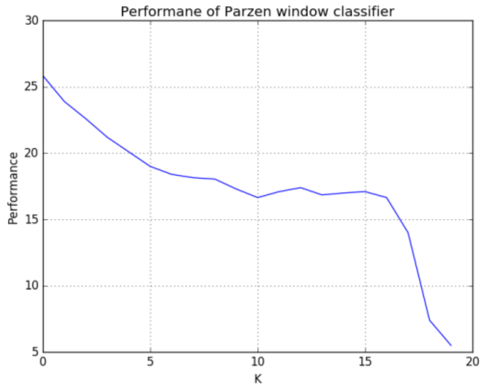

Figure 8: Performance of Parzen window classifier on satimage dataset

4 Conclusion

As shown in figures and numerical result it's obvious that all three classifier perform well classification on synthetic 2d normal distribution. This promising classification is derived from this point that two classes are almost separable as shown in figure 1. It seems that *k*NN classifier is a powerful classifier related to the others, because this classifier has higher percent of performance on the datasets. By analyzing figures more precisely we think there is an optimum value for k to achieve more discriminated classes. By comparing *k*NN and Parzen window classifier it seems that their behaviors on first two datasets are almost equal except the last one (satimage) that *k*NN do better.

There are advantages and limitations for classifiers, in *k*NN classes need not to be linearly separable and no training step needed, in versus computational time to find nearest neighbor increases by increasing datasets samples, in the other word performance of algorithm depends on the number of dimensions used. For bayes classifier, it's simple to implement, great computational efficiency and the precision of algorithm decreases if the amount of data is less and for obtaining good results it requires a very large number of records. Parzen window classifier can be applied to data from any distribution and in contrast choosing the appropriate window size h is difficult.

References

[1] Fisher, Ronald A. "The use of multiple measurements in taxonomic problems." Annals of eugenics 7.2 (1936): 179-188.
[2] Merz, C., Murphy, P.: UCI repository of machine learning databases (1998) www.ics.uci.edu/mlearn/MLRepository.html
[3] Nikam, Sagar S. "A comparative study of classification techniques in data mining algorithms." *Orient J Comput Sci Technol* 8.1 (2015): 13-19.