IJCSC Volume 4 • Number 2 September 2013 pp.18-20 ISSN-0973-7391

KNN and SVM

Aarti kaushik ,Gurdev Singh
M.tech student, Jind institute of Engineering and Technology, Haryana.
H.O.D, Jind Institute of Engineering and technology.
kaushik.aarti@gmail.com

Abstract- In this paper we discuss the characteristics of some classification methods such as KNN and SVM. We will also discuss the performance comparison and evaluation of SVM and KNN.

INTRODUCTION

E-linear classifier is frequently used in practice. Non Linear classifier serves alternative choices, since they admit non linear boundary and more flexibility. For example:-Nearest neighbor classifier, kernel SVM, tree, forest and so on. Nearest neighbors have been used in statistical estimation and pattern recognition already in the beginning of 1970's.In knn training can be very easy, just memorizing training instances. SVM is generally used for binary classification. SVM are currently best performers for classification tasks ranging from text to genomic data. SVM can also be applied to complex data types beyond feature vectors(e.g. graphs, sequence, relation data) by designing kernel functions from such data. SVM are a form of instance based learning.

1. KNN

K-Nearest Neighbor classifiers are based on closeness. When given an unknown tuple, a k-nearest neighbour classifier searches the pattern space for the k training tuples that are closest to the unfamiliar tuple. The k training tuples are the k "nearest neighbors" of the unknown tuple. Closeness is defined in terms of a distance metric such as Euclidean Distance. Nearest Neighbor classifiers can be extremely slow when classifying test tuples .It suffers from poor accuracy when given noisy or irrelevant attributes. Euclidean Distance can be calculated by –

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

A K nearest neighbour is also known as instance based learning in comparison to model based learning because it is not studying any model. A kNN is effective method but simple for classification and is able to solve complex problems. The training process is basically learning all the training data. To predict a new data point, we found the closest K (a tunable parameter) neighbors from the training set and let them vote for the final prediction. Voting among the data points is used to decide the classification for data point. The different names of KNN are Memory Based reasoning, Instance based learning, Case based Reasoning, Example Based Reasoning, Lazy learning.

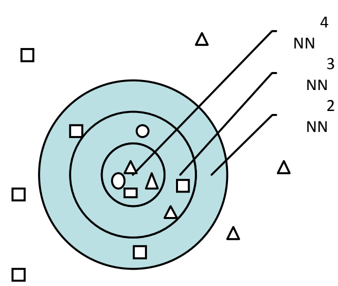

Figure 1:-Example of 4, 3 and 2 nearest neighbors

KNN used in various applications in the field of data mining, statistical pattern recognition and many others.
The major drawbacks of KNN are

(a) Low efficiency
b) Its dependency on the “good value” for k.
c) Cannot handle high number of dimensions well.
D) testing can be very expensive, requiring detailed comparison to all past training instance

I. SUPPORT VECTOR MACHINE

SVM is based on supervised learning which classifies points to one of two disjoint half-spaces. It uses nonlinear mapping to convert the original data into higher dimension.
SVM has two approaches for binary classification

1. Case when the data are linearly separable
2. Case when the data are non-linearly separable

SVM have various properties: Flexibility in choosing a similarity function, feature selection, overfittting can be controlled by soft margin approach, ability to handle large feature space. Support vectors are used to specify the separating hyperplane. For multiclass classification, binary SVM are combined, in either one versus one scheme or one versus rest (OVR) or directed acyclic graph (DAGSVM).In one versus all scheme, Train n binary classifiers ,one for each class against all other classes Predict class is the class of the most confident classifier. SVM implements the structural risk minimization principal which seeks to minimize an upper bound of the generalization error. The main drawback of SVM is training is very slow on complete dataset ,limitation in speed and size during both training and testing phase and selection of kernel functions parameters. In Matlab, Libsvm toolbox is eminent toolbox for SVM classification. There are two categories of SVM that is Linear SVM and Non Linear SVM.
Mathematically, Linear SVM is represented by

$x_i . x_j$

Non Linear SVM map data into new space, then take the inner product of the new vectors. NON Linear SVM is denoted by

$\emptyset(x_i). \emptyset(x_j)$

Kernel function is denoted by

$k(x_i . x_j)$

It means the image of the inner product of the data is the inner product of the images of the data. There are four types of kernel functions: Polynomial and Guassian Radial Basis Function are most important.

A) How to combine SVM with KNN

a) Train a support vector machine on the collection of nearest neighbors
b) Kernel function is defined as:
$K(x,y) = \langle \varphi(x),\varphi(y) \rangle$
Distance function is converted to kernel function
$K(x,y) = \langle x,y \rangle$
$= 1/2(\langle x,x\rangle + \langle y,y\rangle - \langle x-y, x-y \rangle)$
$= ½(d(x,0) + d(y,0) – d(x,y))$

There are also various other ways of transforming a distance function into a kernel function, too. Combining SVM and KNN has various advantages: training an SVM on the entire set is leisurely and the extension to a multiple classes is not familiar as NN. However in the neighborhood of a limited number of examples and a limited number of classes, SVM often implement better performance than other classification methods.

II. PERFORMANCE COMPARISON AND EVALUATION

We compare the characteristics of two kinds of classifiers in the following respects

Training Complexity:-
The parameters of Knn classifiers are generally modified by distance measures .SVM are generally trained by quadratic programming and the training time is proportional to the square of the number of the samples.

Storage and execution complexity:- KNN classifiers have very less parameters and are simple to control. SVM learning by quadratic programming results in large number of support vector is stored and computed in classification. SVM preoccupy high storage and computation than KNN.

Adaptability of training: The parameter of KNN can modify in feature weighting and also simple to add new class to an existing classifier. On the other hand SVM classifier is reciprocal to square of the number of classes and to assure the stability of parameters, adding new classes or new samples need retraining with all samples.

Classification accuracy:-

SVM have been manifested superior classification accuracies to KNN classifiers in many experiments. When training with enough samples, SVM classifiers give higher accuracy than other classifiers.

CONCLUSION
It is obvious that combining SVM and Knn provide better results. Most of the results are even better than the traditional methods like KNN and SVM. There are various issues that are associated with SVM such as choice of kernel, choice of kernel parameters, optimization criteria. Although the performance of KNN was very low as compared to SVM.

References
H Zang, ALexender C.Berg Michael Maire Jitendra Malik, Computer Science division, EECS department, University of California, SVM KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition.

1. S Mohanty, Performance comparision of SVM and K-NN for oriya character, http://thesai.org/Downloads/SpecialIssueNo1/Paper_16-Performance%20Comparison%20of%20SVM%20and%20K-NN%20for%20Oriya%20character%20recognition.pdf
2. Jiawei Han and Micheline Kamber, Data Mining :Concepts and Technique
3. Dr Saed Syan, University of Toronto, Support Vector machines
4. http://cse.seu.edu.cn/people/xgeng/files/under/S08_a.pdf
5. http://www.d.umn.edu/~rmaclin/cs8751/Notes/L06_Instance_Based_Learning.pdf