SVM Kernel Functions for Classification

Arti Patle

Deepak Singh Chouhan

Abstract-A **new generation learning system based on recent advances in statistical learning theory deliver state-of-the-art performance in real-world applications that is Support Vector Machines [2]. Applications such as text categorization, hand-written character recognition, image classification, bio-sequence analysis [9] etc for the classification and regression Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice. In this paper, a novel learning method, Support Vector Machine (SVM), is applied on different data. This paper emphasizes the classification task with Support Vector Machine with different kernel function. It has several kernel functions including linear, polynomial and radial basis for performing classification [13].**

Keyword: **Kernel, radial basis function, feature, support vector.**

I.INTRODUCTION

Computer science has the power to achieve result. In term of result there are many concern either it is database, machine language, robotics, nanotechnology, real-time application [3] etc. The domain, we emphasis is data mining [19]. Data mining domain also divided into number of tasks such as association rule clustering, decision tree and classification. For the classification we focus on Support Vector Machine (SVM). SVM is a powerful state-of-the-art algorithm with strong theoretical foundations based on the Vapnik-Chervonenkis theory. SVM has strong regularization properties. Regularization refers to the generalization of the model to new data. SVMs are a set of related supervised learning methods used for classification and regression [1].

II. SVM

A SVM performs classification by constructing an *N*-dimensional hyperplane [11]. An SVM is a mathematical entity, an algorithm for maximizing a particular mathematical function with respect to a given collection of data. SVM has four basic concepts:

1. The separating hyperplane
2. The maximum-margin hyperplane
3. The soft margin
4. The kernel function [8]

2.1 SEPARATING HYPERPLANE

Straight line divides the space in half, and in three dimensions, we need a plane to divide the space. The general term for a straight line in a high-dimensional space is a hyperplane [17], and so the separating hyperplane is essentially the line that separates the different data set samples (Fig. 1).

2.2 THE MAXIMUM-MARGIN HYPERPLANE

The maximum-margin hyperplane means selecting a line in the middle [18]. In other words, one would select a line that separates the two classes but adopts the maximal distance from any one of the given

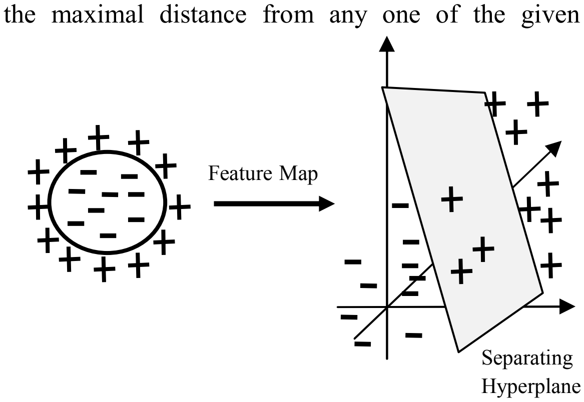

Fig. 1. Separating hyperplane

expression profiles.

2.3 THE SOFT MARGIN

SVM to be able to deal with errors in the data by allowing a few anomalous expression profiles to fall on the 'wrong side' of the separating hyperplane. To handle cases like these, the SVM algorithm has to be modified by adding a 'soft margin' (Fig. 2). Essentially, this allows some data points to push their way through the margin of the separating hyperplane without affecting the final result. In the following Fig. 2 the hyperplanes are B1 and b2. B1 is better than B2 because it maximizes the margin.

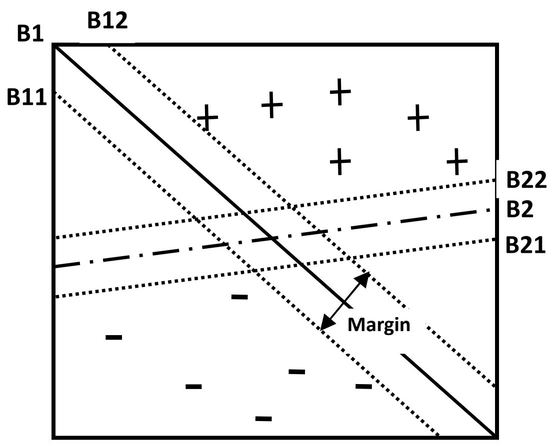

Fig.2. Soft Margin

Given a training set of N data points $\{(x_i, y_i)\}_{N i=1}$, with input data x_i and corresponding binary class labels y_i €{−1,+1}, the SVM classifier, according to Vapnik's original formulation [1] satisfies the following conditions [16]:

$w^T\varphi(x_i) + b \geq +1$, if $y_i = +1$
$w^T\varphi(x_i) + b \leq -1$, if $y_i = -1$

Which is equivalent to

$y_i[w^T \varphi(x_i) + b] \geq 1, i = 1, ...,N$

$$\alpha_i^0 = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Here $K(x_i, x_j)$ is a kernel function

2.4 THE KERNEL FUNCTION

The kernel function is a mathematical trick that allows the SVM to perform a 'two-dimensional' classification of a set of originally one-dimensional data. In general, a kernel function projects data from a low-dimensional space to a space of higher dimension.

$$\langle x_1 \cdot x_2 \rangle \leftarrow K(x_1, x_2) = \langle \Phi(x_1) \cdot \Phi(x_2) \rangle$$

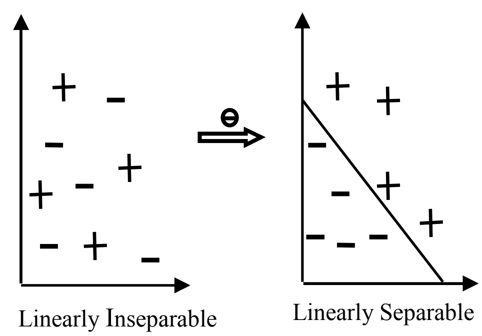

Fig.3. Kernel Function Mapping

2.4.1Linear Kernel Function: Linear kernel function is commonly described as

$$K(x, x_i) = x \cdot x^T$$

2.4.2 Polynomial Kernel Function: The polynomial kernel function is directional, i.e. the output depends on the direction of the two vectors in low-dimensional space. This is due to the dot product in the kernel. The magnitude of the output is also dependent on the magnitude of the vector x_i

$$K(x, x_i) = \left(1 + x \cdot x_i^T\right)^d$$

'd' is degree of kernel function.

2.4.3 Radial Basis Function: Radial basis function is one of the most popular kernel functions. It adds a"bump" around each data point.

$$K(x, x_i) = e^{-\gamma \| x - x_i \|^2},$$

kernel function parameter $\gamma > 0$

2.4.4 Sigmoid Function:

$$K(x_i, x_j) = \tanh\left(\gamma x_i^T x_j + r\right)$$

Here γ, r and d are kernel parameters.

Selection of kernel function is depends on the application. It is not fixed.

III. WHY SVM PREFER

1. near-perfect classification accuracy on the data set [9]
2. SVM can deal with more than two predictor variables.
3. Separating the points with non-linear curves.
4. Handling the cases where clusters cannot be completely separated.
5. Handling classifications with more than two categories.
6. SVM has the ability to handle large number of attributes (up to 300000) of a dataset

IV. **APPLICATION & LIMITATION OF SVM**

In Data mining SVM deals Classification and regression problem [19]. SVM application domain is wide and relate with real world problem such as:

1. SVM for Credit scoring [2].
2. SVM for Geo and Environmental Sciences [4].
3. SVM for Protein Fold and Remote Homology Detection [5].
4. Content based image retrieval [6].
5. Facial expression classification [7].
6. SVMs for Texture Classification.
7. SVM application in E-learning.
8. NewsRec, a SVM-driven Personal Recommendation System for News Websites [10].
9. A Comparison of the Performance of Artificial.
10. Neural Networks and SVMs for the Prediction of Traffic Speed and Travel Time [11].
11. Protein Structure Prediction[12].
12. Face Reorganization with SVM [20].

4.1 Limitations of SVM can be:

1. Most notable limitation of the SVM lies in the choice of the kernel function.
2. Second limitation is speed and size (millions), both in training and testing.
3. Third limitation is the lack of efficiency in handling multiclass problem. Multiclass classifier is an area of current research.

V.**FEATURE SELECTIONS**

The classification accuracy rate of SVM is influenced by not only the two parameters C and γ, but also other factors including the quality of the feature's dataset. For instance, the correlation between features influences the classification result [15]. Accidental elimination of important features might decrease the accuracy rate of classification. Additionally, some features of the dataset may have no effect at all, or contain a high level of noise [17]. Removal of such features can increase the search speed and the accuracy rate. Feature selection is a part of supervised classification [14] .two approach for feature selection.

5.1 filter-based feature selection

The filter approach selects important features first and then SVM is applied for classification [15].

5.2 wrapper-based feature selection

SVM to choose important features as well as conducts training/testing or combines SVM with other optimization tools to perform feature selection.

Here we assume the selected feature for the calculation purpose. (Consider for credit scoring [2])

The accuracy of machine learning product are highly depends on training. Suppose a machine was trains for 100 numbers finding even odd number than in case of testing numbers are range lies 100 than the result surety is good If the input was higher number than might be result was not accurate.

VI. **PROCEDURE**

Step1. Select the dataset which has multiple instances.

Step2. Select attributes, this study use the selected attributes.

Step3. Set the SVM parameter C=1.0, γ=0.01, d=2

Step4. Select kernel for classification, Polynomial or RBF.

Step5. Comparisons between classified instances and unclassified instances.

Fig 4 & Fig 5 shows the comparison with RBF & Polynomial kernel function.

DataSet:
In this study use the dataset in ARFF format. ARFF give the detail description of dataset with attributes name, relation name, class variable. As following dataset [24]:

Relation name: Credit
Attribute Name: A1…. A15
Attributes values lies in {}

```
@RELATION credit
@ATTRIBUTE A1 {b,a}
@ATTRIBUTE A2 REAL
@ATTRIBUTE A3 REAL
@ATTRIBUTE A4 {u,y,l,t}
@ATTRIBUTE A5 {g,p,gg}
@ATTRIBUTE A6 {c,d,cc,i,j,k,m,r,q,w,x,e,aa,ff}
@ATTRIBUTE A7 {v,h,bb,j,n,z,dd,ff,o}
@ATTRIBUTE A8 REAL
@ATTRIBUTE A9 {t,f}
@ATTRIBUTE A10 {t,f}
@ATTRIBUTE A11 REAL
@ATTRIBUTE A12 {t,f}
@ATTRIBUTE A13 {g,p,s}
@ATTRIBUTE A14 REAL
@ATTRIBUTE A15 REAL
@ATTRIBUTE class {+, -}
@DATA
b,30.83,0,u,g,w,v,1.25,t,t,01,f,g,00202,0,+
a,58.67,4.46,u,g,q,h,3.04,t,t,06,f,g,00043,560,+
a,24.50,0.5,u,g,q,h,1.5,t,f,0,f,g,00280,824,+
b,27.25,0.625,u,g,aa,v,0.455,t,f,0,t,g,00200,0,-
b,37.17,4,u,g,c,bb,5,t,f,0,t,s,00280,0,-
b,?,0.375,u,g,d,v,0.875,t,f,0,t,s,00928,0,-
```

VII. CALCULATION EVALUTION

As describe the mathematical logic in chapter 2, in this evaluation map the calculation with dataset. Datasets are in arff (attribute relation file format) .

- **Polynomial kernel Function:**

In this example, classification of even and odd numbers is solved by SVM 2^{nd} degree polynomial kernel. This is a non-linear classification problem.

```
@RELATION evenodd
@ATTRIBUTE A1 REAL
@ATTRIBUTE class {+,-}
@DATA
2 , +
4 , +
1 , -1
```

Input Data x	Output Class y
2	1
4	1
1	-1

First, we transform the dataset by polynomial kernel as:

$$K(x^i, x^j) = (1+x^{iT} x^j)^2$$

Here, $x_i^T x_j = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} * [\,2 \quad 4 \quad 1\,]$

Now the kernel matrix is :

$$K(x_i, x_j) = \begin{bmatrix} 25 & 81 & 9 \\ 81 & 289 & 25 \\ 9 & 25 & 4 \end{bmatrix}$$

Now, $\alpha_i^0 = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$

$$\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left(25\alpha_1^2 + 81\alpha_1\alpha_2 + 9\alpha_1\alpha_3 + 81\alpha_2\alpha_1 + 289\alpha_2^2 + 25\alpha_2\alpha_3 + 9\alpha_3\alpha_1 + 25\alpha_3\alpha_2 + 4\alpha_3^2\right)$$

$$= \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left(25\alpha_1^2 + 162\alpha_1\alpha_2 - 18\alpha_1\alpha_3 + 289\alpha_2^2 - 50\alpha_2\alpha_3 + 4\alpha_3^2\right)$$

Differentiating with respect to Lagrangian parameter $\{\alpha_1, \ldots, \alpha_3\}$ we obtain the following set of simultaneous equation:

$$25\alpha_1 + 81\alpha_2 - 9\alpha_3 = 1$$
$$81\alpha_1 + 289\alpha_2 - 25\alpha_3 = 1$$
$$-9\alpha_1 - 25\alpha_2 + 4\alpha_3 = 1$$

By solving the above equation we get,

$$\alpha_1 = 14$$
$$\alpha_2 = 15.77$$
$$\alpha_3 = -2.55$$

Therefore, the decision function is:

$$F(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x)$$

Since the kernel considers the inner product of input vectors, we can write the 2nd degree kernel function as

$$K(x_i, x_j) = ((x_i, x_j) + 1)^2$$

$$= \left(x_{i1}x_{j1} + x_{i2}x_{j2}\right)^2 + 2\left(x_{i1}x_{j1} + x_{i2}x_{j2}\right) + 1$$

$$= 1 + \left(x_{i1}x_{j1}\right)^2 + 2\left(x_{i1}x_{j1}\right)\left(x_{i2}x_{j2}\right) + \left(x_{i2}x_{j2}\right)^2 + 2\left(x_{i1}x_{j1}\right)$$

$$= \Phi\left(x_i\right)^T \left(x_j\right)$$

The second degree polynomial transformation function can be written as

$$\Phi\left(x_i\right) = \left[1, x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}\right]^T$$

Input Space	Feature			Output
x	1	x_i^2	$\sqrt{2}x_i$	
2	1	4	2.828	1
4	1	16	5.656	1
1	1	1	1.414	-1

By substituting the value of α and $\Phi(x_i)$ in the equation below, we calculate the value of ω^0

$$\omega^0 = \sum_{i=1}^{3} \alpha_i y_i \Phi\left(x_i\right)$$

$$= 14 \begin{bmatrix} 1 \\ 4 \\ 2.828 \end{bmatrix} - 2.55 * \begin{bmatrix} 1 \\ 16 \\ 5.656 \end{bmatrix} - 15.77 \begin{bmatrix} 1 \\ 1 \\ 1.414 \end{bmatrix}$$

$$= \begin{bmatrix} -4.32 \\ -0.57 \\ 2.867 \end{bmatrix}$$

Now finally the hyperplane can be defined as $\left(\omega^0\right)^T \Phi(x) = 0$ which translate to

$$\begin{bmatrix} -4.32 & -0.57 & 2.867 \end{bmatrix} * \begin{bmatrix} 1 \\ x^2 \\ \sqrt{2}x \end{bmatrix}$$

$$.57x^2 - 4.054x + 4.32 = 0$$

If take i/p 3 then F(3) is -1 or odd.

VIII. EMPIRICAL ANALYSIS

The choice of kernel is an important issue in the SVM algorithm, and the performance of it largely depends on the kernel. Up to now, no general rule is available as to which kernel should be used. In this paper we investigate two kernels: RBF kernel and polynomial kernel. So far RBF kernel is the best choice for practical applications.

In this paper we use the data sets which are available in the UCI Repository of Machine Learning Databases [21] and are adopted herein to evaluate the Predictive accuracy. The dataset which has different number of attributes and number of instances; i.e. credit dataset, generated credit dataset, titanic dataset, vehicle dataset, train dataset.

Table. 1 shows the different dataset name, number of instances in each dataset , kernel function name from which classification task perform, percentage of correctly classified and incorrectly classified instances. The experimental result shows that when the dataset is small than the accuracy not differs but when the dataset is large the RBF kernel is more accurate than polynomial kernel function for classification task of dataset. Fig. 4 and fig. 5 are Comparison Graph b/w RBF & Polynomial Kernel for correctly classified and incorrectly classified instances in Different Dataset.

TABLE1: Result summary with polynomial and RBF kernel for different DATASETS (DS).

NAME OF DATA SET		CREDIT DS	GENERATED CREDIT DS	MARKS DS	TITANIC DS	VOTE DS	VEHICLE DS	VOWEL DS	TRAIN DS
KERNEL NAME / NO. OF INSTANCES		490	50	5	2201	435	846	435	10
RBF	Correctly Classified Instances	55.71%	58%	60%	67.7%	96.09%	74.35%	96.09%	70%
	Incorrectly Classified Instances	44.28%	42%	40%	32.30%	3.90%	25.65%	3.90%	30%
POLYNOMIAL	Correctly Classified Instances	86.35%	58%	60%	77.6%	94.48%	40.07%	94.48%	10%
	Incorrectly Classified Instances	13.4694 %	42%	40%	22.39%	5.5172%	59.92%	5.51%	90%

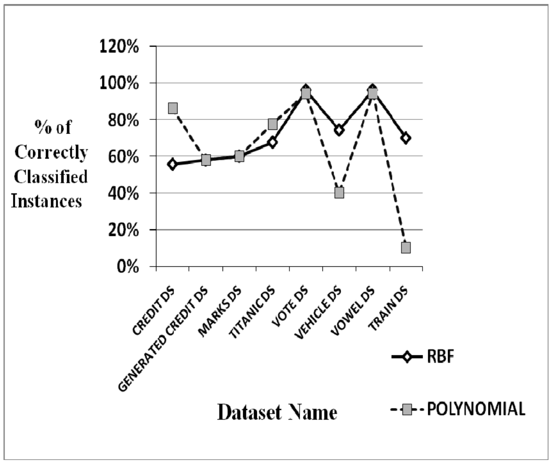

Fig.4. Correctly classified instances

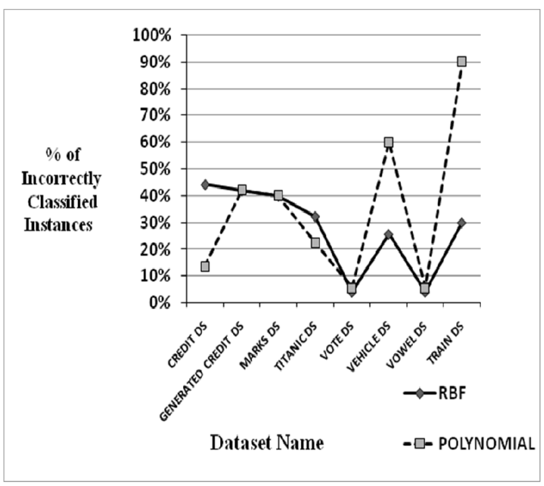

Fig.5. Incorrectly classified instances

IX. CONCLUSION

SVM is an emerging data classification technique first developed by Vapnik [1], and has been widely adopted in various fields of classification problems recently. The SVM has been introduced as a robust tool for many aspects of data mining including classification, regression and outlier detection. Rule extraction techniques generate classification models that have clear advantages. First of all, they are comprehensible and therefore easy to incorporate in real-life applications where clarity of the classifications made is needed. Secondly, the extracted rules only lose a small percentage in accuracy of the black box model from which they are generated. Since SVMs are among the best performing classifiers, rules extracted from SVMs achieve an accuracy that often surpasses that of the classical methods, such as C4.5 and logit. Kernel selection is the main task in SVM. Through kernel selection we find the comparison of accuracy between Dataset. Empirical analysis of kernel selection shows that kernel selection reflects the accuracy of classification.

X. REFERENCES

[1] Wikipedia Online. Http:// en. wikipedia. org/wiki

[2] Cheng-Lung Huang , Mu-Chen Chen, Chieh-Jen Wang , Credit scoring with a data mining approach based on SVMs, National Kaohsiung First university of Science and Technology, Department of Information Management Nantz District, Kaohsiung 811, Taiwan, pp. 847-856, 2007

[3] Serdar Iplikci, Support vector machines-based generalized predictive control, , INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL,Vol.16, pp. 843-862,2006 http://ietfec.oxfordjournals.org/cgi/content/abstract/E89-A/10/2787

[4] N. Gilardi, M. Kanevski, M. Maignan and E. Mayoraz. Environmental and Pollution Spatial Data Classification with Support Vector Machines and Geostatistics. Workshop W07 Intelligent techniques for Spatio-Temporal Data Analysis in Environmental Applications. ACAI99, Greece, July, 1999. pp. 43-51.www.idiap.ch

[5] Huzefa Rangwala and George Karypis Profile based direct kernels for remote homology detection and fold recognition Bioinformatics, vol-21, pp. 6-22, 2005

[6] Dacheng Tao, Xiaoou Tang, Xuelong Li, and Xindong Wu, Asymmetric Bagging and Random Subspacing for Support Vector Machines-based Relevance Feedback in Image Retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence , vol-28, pp. 1088 – 1099, 2006.

[7] J. Ghent and J. McDonald, Facial Expression Classification using a One-Against-All Support Vector Machine, proceedings of the Irish Machine Vision and Image Processing Conference, Aug 2005.

[8] Nello Cristianini & John Shawe-Taylor, An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, chap-6, 2000 http://www.supportvector.net

[9] Furey, T.S. et al, Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16, pp. 906–914 (2000).

[10] Christian Bomhardt, NewsRec, A SVM-driven Personal Recommendation System for News Web Intelligence, IEEE/WIC/ACM International Conference on (WI'04), pp.545-548, 2004.

[11] S.R.Gunn, Support Vector Machines for Classification and Regression, http:// www. ecs. Soton .ac.uk /~srg/ publications/ pdf/SVM.pdf, pp. 5-28, 1998.

[12] Kim, H. and H. Park, Prediction of protein relative solvent accessibility with support vector machine and long range 3D local descriptor, pp. 557-62, Proteins 2004.

[13] Nello Cristianini and John Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000, .

[14] Tony Bellotti and Jonathan Crook , Support vector machines for credit scoring and discovery of significant features, Credit Research Centre School of Management and Economics University of Edinburgh, 7 May 2007.

[15] Chen, Y.-W., & Lin, C.-J. , Combining SVMs with various feature eselection strategies. http: // www .csie .ntu .edu .tw/ ~cjlin / Papers/ features.pdf 2005.

[16] C. Cortes and V. Vapnik. Support vector networks. Machine Learning, 20, pp. 273 – 297, 1995

[17] Brill, J.,The importance of credit scoring models in improving cash flow and collection. Business Credit, 100(1), pp. 16–17, 1998.

[18] Chang, C. C., & Lin, C. J. LIBSVM: a library for support vector machines. http: // www. csie. ntu. edu.tw/~cjlin/libsvm, 2001.

[19] Chen, S. Y., & Liu, X. The contribution of data mining to information science. Journal of Information Science, 30(6), 550–558, 2004.

[20] Bernd Heisele_ Purdy Ho_ Tomaso Poggio, Face reorganization with support vector machine Global versus component based approach, July 2001.

[21] http://repository.seasr.org/Datasets/UCI/arff/

AUTHOR'S INFORMATION

Arti Patle
Department of Computer Science & Engineering,
IMS Engineering College, Gaziabad (UP)
artipatle@gmail.com

Deepak Singh Chouhan
Department of Computer Science & Engineering
IES, IPS Academy , Indore(MP)
Deepak.ur@gmail.com