

---

# IMPLEMENTATION OF NEURAL NETWORK MODELS ON BUDDY MOVE DATA SET

---

Using Perceptron, SVM, SOM and LVQ



OCTOBER 2, 2020  
LOVELY PROFESSIONAL UNIVERSITY  
Phagwara, Punjab

# Implementation of Neural Network Models on Buddy Move Data Set

Nitin Sanatan

Student, B.Tech. (Computer Science and Engineering)

Lovely Professional University, Punjab, India

## Abstract

Humans and Animals have ability to continuously acquire and transfer knowledge throughout their lifespan & for this task only, we are known as intelligent systems. From mid 1950s, scientists like Warren McCulloch, Walter Pitts and D.O Hebb had opened the subject of computational neural network. Since then, many advancements had happened like creation of perceptron by Rosenblatt (1958), functional multilayer network were published by Lapa (1965) and continuous backpropagation using Dynamic Programming (1961). ANNs began as an attempt to duplicate the structure of human brain to perform tasks that conventional algorithms could do. ANN replicate the brain using artificial neurons, use learning and predict the result like what a human brain does. Nowadays, ANN is used in many industries for function approximation or regression analysis, data processing and classification purpose. Its ability to process nonlinear model is extensively used in System Identification and control (vehicle control, trajectory prediction), image processing, pattern and sequence recognition (handwriting) and 3D reconstruction. ANNs have been used to accelerate reliability analysis of infrastructures subject to natural disasters and to predict foundation settlements. ANNs have been employed in cybersecurity, with the objective to discriminate between legitimate activities and malicious ones. For example, machine learning has been used for classifying Android malware, for identifying domains belonging to threat actors and for detecting URLs posing a security risk.

## Keywords

Artificial Neural Network (ANN), Buddy Move Dataset, Classification, Clustering, Supervised Learning, Unsupervised Learning.

## Introduction

Human brain is the most advanced and intelligent among all the creatures. It generally recognizes the pattern of everything and make some inferences based on it. Nowadays, Machines are also able to recognize the patterns in a dataset and approximate the results using different learning techniques.

One of the best features of Machine Learning Algorithms are, they can be trained both on labelled (Supervised Learning) and unlabeled data (Unsupervised Learning). Unsupervised learning is a technique of model grouping the data without knowing what the actual objects are.

Neural Networks are highly efficient in classification (a supervised technique), clustering (unsupervised techniques). It can also be trained to process, identify and predict the image using Convolutional Neural Network. For classification and clustering, we have different algorithms in Neural Network which can determine the classes and clusters based on training data. From a structural perspective, ANNs can be divided into two main categories, feed-forward networks, in which the computation is performed in a layer-by-layer fashion from the input to the output of the network; and recurrent networks which have an interconnected network structure including cycles.

All these neural networks take a fixed amount of input and output data to be trained, in a fixed format (array, data frame etc.). Before the data is fed to network, weights are randomly initialized corresponding to each neuron. These weights are further updated based on the error (difference between actual output and obtained output) and these weights are again fed to neurons and iteration continues, to get the most appropriate results. Once the data is fed, our network is trained by iterative learning method. We use these ANN algorithms in different dataset to classify and clustering of data. For Example, in *IRIS* dataset, the features of a flowers are used to distinguish the flower. *Iris setosa*, *Iris virginica* and *Iris versicolor* are the features used in *IRIS* dataset.

## Artificial Neural Network on Buddy move dataset

The main motive of this project is to use ANN techniques and predict several trends based on the reviews posted on *holidayiq.com*. We have Buddy move dataset which contains the numbers of reviews information in every category.

We need to predict the number of reviews given in a category based on reviews given on other categories. We are mainly focused upon classifying the user interests based on number of reviews published by them on the website.

In ANN, I have used Perceptron, Backpropagation, Support Vector Machine, Self-Organizing Maps and Linear Vector Quantization for training my data.

In Supervised learning like Perceptron, Backpropagation, SVM and LVQ techniques, we have considered the data or features as 'all attributes except Sports' and 'Sports' attribute is our target data. We are basically trying to classify user interests through their point of interest ('Shopping Malls', 'Picnic', 'Theatre', 'Nature' etc.)

In the Unsupervised learning, Self-Organizing Map, we have tried to cluster the users based on their interests. There are basically 2 clusters, one which contains the people more inclined towards sports/fitness and the other cluster contains those who are fun-loving (doesn't like sports much)

In our case, Neural Network takes input of dimension 5. There are total 249 rows in our dataset, so data is of shape (249,5) and target is of shape (249,).

## **Related Work**

Artificial neural network is a computational model that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions. From the past many years, there has been a lot of advancements in the field of Neural Networks. Classification and Clustering algorithms have been continuously improving and being efficient in predicting results. Due to its reliable performance, neural networks applications has risen from past years from banking, finance to environment and geology.

In year 2016, Y.S. Park and S. Lek, the practical application of Multilayer Perceptron in ecological model was presented. In 2019, C. Gajendran used Bayesian probability network models to predict the water quality of Nambiyar River in Tamilnadu, India. In the same year, Syed Muzamil and Dharmendra Singh Rajput published their research work in Deep Learning and Parallel Computing Environment in Bioengineering Systems. An SVM was used in developing a computer-aided diagnosis (CAD) system which can perform analysis on MR images, helping in medical interpretation.

## Literature Survey

Author	Year	Title	Literature Review
D.E. Rumelhalt, D. Zipser	1985	Feature Discovery by Competitive Learning	Each cluster classifies the stimulus set of M groups, one for each unit in the cluster. Update weight if the unit wins(active).
Herbert L. Jensen	1992	Using Neural Networks for Credit Scoring	Utilization of ANN in credit scoring using personal DOS computer.
Kate A. Smith, Jatinder N.D. Gupta	2002	Neural Networks in Business: Techniques and Applications	Authors in this book has explained the application of ANN in all segments of business
Rui Xu	2005	Survey on Clustering Algorithms	Implemented Clustering algorithms like Self-Organizing feature map (SOFM) on benchmark data sets, travelling salesman problem and bioinformatics.
Emny Harna Yossy	2017	Mango fruit sortation system using Neural Network and Computer Vision	Network consisted 40 hidden layers. ANN predicts the color and size of mango with 94% accuracy.

## Dataset

Dataset plays the most important role in any machine learning algorithms. Just like humans needs their eyes and sense organs to observe the things happening around and recognizing the pattern, similarly dataset is fed to networks and machine recognizes the pattern. In my work, buddy move dataset is collected from UCI Machine Learning website. It contains the number of reviews posted by users on *holidayiq.com* and a total of 249 data are there in the Buddy Move dataset. The first attribute 'Sports' is taken as Target and rest all attributes 'Shopping Malls', 'Theatre', 'Picnic', 'Nature' is taken as training data.

## Limitations of Artificial Neural Network

- **Black Box Nature:** When it comes to output then Neural Network predicts the output with a great accuracy but the process going in networks (hidden layers) are still unclear. We are still not able to understand by which means it is concluding or inference.
- **Duration of development:** Although libraries like 'keras' and 'tensorflow' makes the computational task a lot easier still there exists a major drawback in duration required by neural network to build and train the data. It's easier to work on the task which has already been done but it's a lot difficult to make efficient algorithms when we are solving a new problem in machine learning.
- **Amount of Data:** ANN require a very large data set to train our model and obtain great accuracy in results. With few algorithms like Naïve Bayes, network works well with small dataset too. But, in most of the cases, we require a good amount of labelled/unlabeled data for building and training our model.
- **Computationally Expensive:** Neural networks are also more computationally expensive than traditional algorithms. As neural networks require much larger dataset, it also require more time to train the data and similarly takes larger time for computations.

## Proposed Methodology

For classification and clustering of Buddy move dataset, I have implemented different neural networks algorithms like Perceptron [Figure 1], MLP using Backpropagation, SVM [Figure 2], SOM [Figure 3] and LVQ.

Data is given to the input of neural network then neurons adjust weights corresponding to every neuron (if any error is identified), the most appropriate result is taken as output after comparing it to the target dataset.

In Supervised learning like Perceptron, LVQ etc. algorithms continue with weight update.

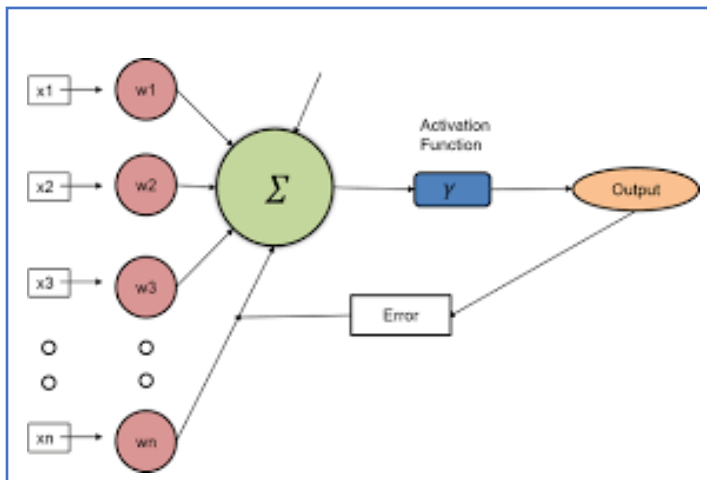


Figure 1: Perceptron Model

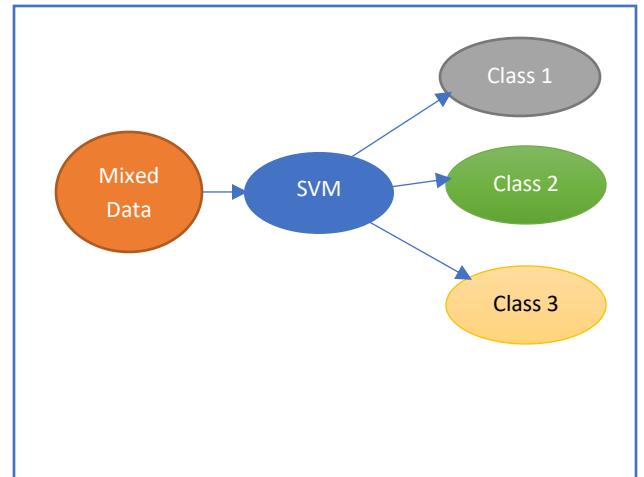


Figure 2: Support Vector Machine

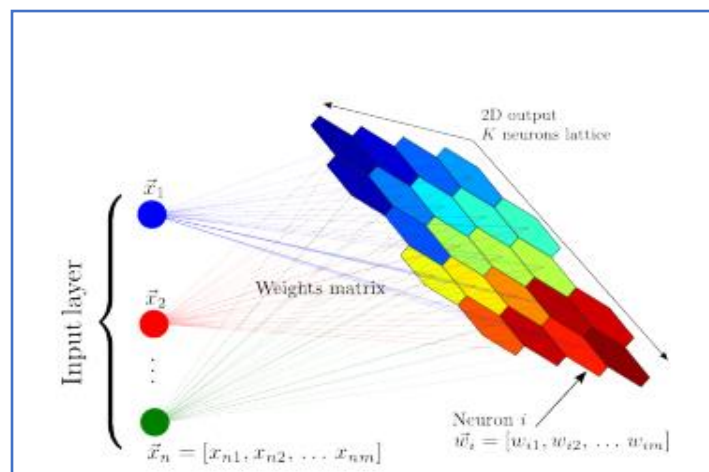


Figure 3: Self Organizing Map (Unsupervised)

## Results and Discussion

Buddy move dataset is trained using different neural network techniques. I have used Google Colab platform to implement my code. I have used libraries like keras[1], tensorflow[1] and Minisom[2] to implement MLP using Backpropagation[1] and Self-Organizing Maps[2].

Also, for Support Vector Machine and Perceptron, I have used svm and sklearn library.

In classification using MLP (backpropagation), the accuracy obtained was very low because it requires a large dataset and, in our case, dataset contains just 249 entries. Accuracy occurred using perceptron is also quite low(15%) because of the small dataset and since the values (number of reviews published) are very variable.

In Support Vector Machine, I obtained a 60% accuracy on testing data. The model worked so perfect because I have made data linear separable using kernel trick. Also, I have obtained the support vector indices in order to track those vectors lying on the classifier boundary.

In Self-Organizing map and Linear Vector Quantization, weight are getting updated and learning rate parameter decreases to 80% after every iteration. The results obtained using these two algorithms are also accurate.

In the following obtained Self Organizing Map [Figure 4], all the attributes except 'Sports' are mapped. We can see the indices neighboring to index (1,10) are those who are most probable to like Sports and Fitness more.

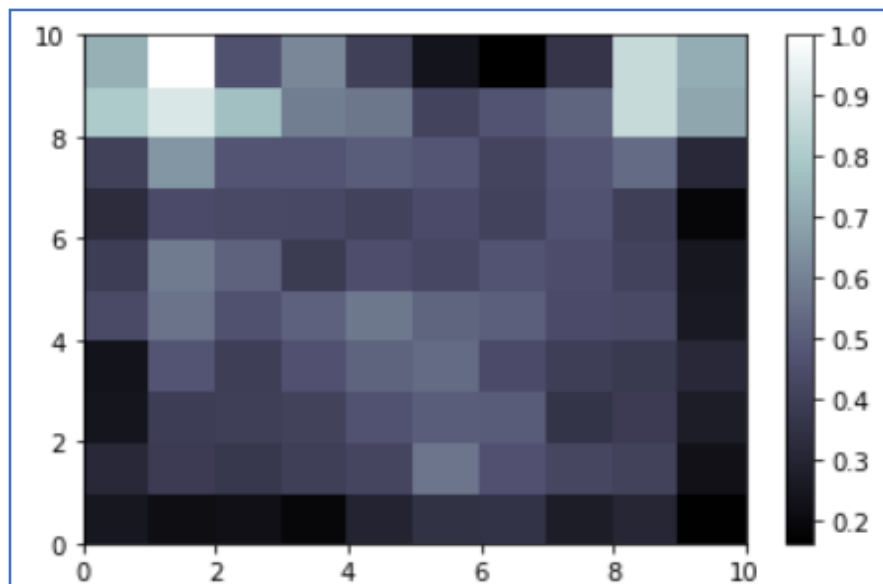


Figure 4: Self Organizing Map on Buddy Move dataset



## Conclusion

In this paper, implementation of ANN on buddy move dataset is discussed. Its evident that ANN are very efficient and accurate in recognizing the pattern of random dataset and classify/cluster data accordingly. ANN techniques outperform traditional computational techniques but not in every case. In case of smaller dataset, the accuracy obtained is low. From above results, it is evident that maximum accuracy is obtained using Support vector machine, which is 80 percent on training data and 60 percent in testing data.

Though ANN cannot contain sequential data like RNN, neither it can identify spatial relationship like CNN but still, ANN are still used for effective forecasting and prediction in many fields.

## References

- [1] S. Renjith and C. Anjali, 'A personalized mobile travel recommender system using hybrid algorithm,' 2014 First International Conference on Computational Systems and Communications (ICCSC), Trivandrum, 2014, pp. 12–17. doi: 10.1109/COMPSC.2014.7032612
- [2] Feature Discovery by Competitive Learning, D.E. Rumelhalt and D. Zipser – 1985  
Cognitive Science, 1985, 9, 75-112, 1985
- [3] McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*. **5** (4): 115–133. doi:[10.1007/BF02478259](https://doi.org/10.1007/BF02478259)
- [4] Hebb, Donald (1949). *The Organization of Behavior*. New York: Wiley. ISBN [978-1-135-63190-1](https://www.isbn-international.org/number/978-1-135-63190-1)
- [5] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model for Information Storage and Organization In The Brain". *Psychological Review*. **65** (6): 386–408

- [6] J. Weng, N. Ahuja and T. S. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," *Proc. International Joint Conference on Neural Networks*, Baltimore, Maryland, vol I, pp. 576–581, June, 1992.
- [7] J. Weng, N. Ahuja and T. S. Huang, "Learning recognition and segmentation of 3-D objects from 2-D images," *Proc. 4th International Conf. Computer Vision*, Berlin, Germany, pp. 121–128, May, 1993.
- [8] J. Weng, N. Ahuja and T. S. Huang, "Learning recognition and segmentation using the Cresceptron," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 105–139, Nov. 1997
- [9] Derrick Mwit, "Introduction to Self-Organizing Maps", heartbeat.com, July 2018
- [10] Y.-S. Park, S. Lek Ecological Model Types in Developments in Environmental Modelling, 2016
- [11] C. Gajendran, P. Thamarai, "Intelligent Prediction Modelling of Water Quality using ANN" in GIS and Geostatistical Techniques for Groundwater Science, Tamilnadu, 2019
- [12] Studies from Department of Computer Science & Engineering in the Area of Artificial Neural Networks Described (A Technique to Stock Market Prediction Using Fuzzy Clustering and Artificial Neural Networks), Atlanta, 2015
- [13] Emny Harna Yossy in "Mango Fruit Sortation System using NN and Computer Vision" et al. 2017
- [14] Koklu, Murat, Tutuncu, K PY - 2019/11/15, SP – 012032, T1 - Tree based classification methods for occupancy detection, VL – 675, DO - 10.1088/1757-899X/675/1/012032, IOP Conference Series: Materials Science and Engineering
- [15] Jay Shah, Neural Networks for Beginners: Popular Types and Applications, 'blogs.statsbot.com', Nov 2017