Classification based on Neural Network (December 2013)

Nikhil Mahajan, Department of Electrical & Computer Engineering, Stevens Institute of Technology, Hoboken, NJ.

Abstract — **Classification pioneers decisive role to make sense of the blooming uncertainty of sensory data that intelligent system confronts. Pattern Recognition foundation can be traced back to the old world, with reference to philosophical epistemology, the study of nature of knowledge. Today, the associated techniques viz. supervised (discrimination) and unsupervised classification (sometimes in statistics, referred to as clustering) are applied in virtually every scientific and technical discipline. In this work, neural network approach has been carried out for diagnosis of breast cancer, to not only provide highly accurate and precise diagnoses that are much less evasive than surgical biopsy but also carry much quicker diagnoses.**

***Keywords* — Artificial Neural Network, ANN, Neural Networks, Feed Forward Network, Back Propagation, Cost function, Regularization.**

I. INTRODUCTION

HUMAN BRAIN is a database where incoming sensory information compares past experiences or learning stored in neurological patterns, that evolves with time and vary spontaneously. This reorganization of the visual information stimulating coherent unfolding of events from past reviews the multisensory perception of senses, sequences assessing the notions of one's internal representation of surroundings or perceptual outcome is seemingly based on the most likely ordering of events given the implicit 'belief system' of the perceptual system. Human tendency to intercept these visions varying from faces, figures in shadows, clouds or pattern with no deliberate design such as swirls on a baked confection has been the building blocks of 'Pattern Classification'. The process concerns with design of systems that detects trends & classify patterns. Patterns being the essence like heart in human body, as being described *essentially an arrangement by N. Wiener*[1].

In machine learning, pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to do ''fuzzy" matching of inputs. Many common pattern recognition algorithms are *probabilistic* in nature, as they implement statistical inference to find the best label for a given instance.

[1] R. C. Gonzalez, "Object Recognition," in *Digital Image processing*, 3rd ed. Pearson, August 2008, pp. 861-909.

II. LITERATURE REVIEW

Neural network is a powerful tool to analyze and evaluate broad spectrum of heterogenous clinical experiments. Most of the applications exercises classification; that is, the task is on the basis of the measured features to assign the patient to one of a small set of classes[2]. Although the predictive accuracy of neural networks is often higher than that of other methods or human experts, it is generally challenging to figure out neural network conclusions beacuse of it's black box nature. This work evaluates whether a neural network trained on a large prospectively collected dataset of consecutive mammography findings can discriminate between benign & malignant disease and accurately predict the probability of breast cancer for individual patients.

A. Breast Cancer Prognosis

Only skin cancer accounts for more diagnoses among women than breast cancer[3]. As with most cancers, early diagnosis of the disease radically increases survivorship. Once a patient is diagnosed with breast cancer the prognosis gives the long term behavior of the ailment. The prognosis is the principal factor in determining the treatment immediately followed by the diagnosis of the disease.

The *Wisconsin Breast Cancer (Diagnostic) Data*[4] provided by the machine learning repository, University of California, Irvine has features extracted from a digitized image of a fine needle aspirate (FNA) of a breast mass which describe characteristics of the cell nuclei present in the image. With number of instances ranging to 569, (357 benign and 212 malignant), each one represents FNA test measurements for one diagnosis case. Each instance has 32 attributes where the first two attributes correspond to a unique identification number and the diagnosis status (benign / malignant). The rest 30 features computes 10 real values feature viz, radius, texture, perimeter, area, symmetry, smoothnes, concavity etc.

[2] . R. Dybowski and V. Gant, Clinical Applications of Artificial Neural Networks, Cambridge University Press, 2007.

[3] Parker S. L. et al, "Cancer Statistics", 1997 *CA-A Cancer Journal for Clinicians*, 47:5-27, 1997.

[4] http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

B. Biological Neural Networks

The human brain is a complex computing system capable of thinking, remembering and solving intensive & exhaustive problems. A *neuron* (or nerve cell) is a special biological cell that processes information (see Figure 1) and compose cell body or *soma* along with two out-reaching trees like brances; the *axon* and the *dendrites*. Soma contains information related to hereditary traits and enclosed plasama in the nucleus holds molecular equipment which generates material needed for neurons to receive signals (impulses) from other neurons through its dendrites (recievers). Signals are transmitted by the cell body along the axon (transmitter), which branchs out into strands and substrands. *Synapses,* an elementary structure and fucntional unit between two neurons (an axon strand of one neuron and dendrite of another) is enclosed at the terminals of strands. Impulses enclosing the synapses releases a certain chemical called neurotransmitters which diffuse across synaptic gap, to enhance or inhibit receptor neuron's own penchant to emit electrial impulses. Efficiency of the signals passing through it can be learned from history of activaties they are involved which is a possible reason for human memory.

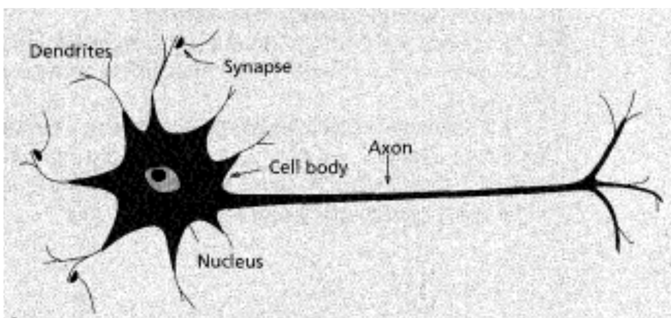

Figure 1. Sketch of a biological neuron.

Massively connected neurons, are complex and dense like telephone networks, for example, cerebral cortex contains about 10^{11} neurons which is approximately the number of stars in the Milky Way[5].

The complex perceptual decision such as face recognition typically carried out in few hundred seconds, network of neurons carry out in few milliseconds.The underlying concept is interaction through a dense web of interconnections unlike conventional computer architectures.

C. Artificial Neural Networks

According to *Dr. Robert Hecht-Nielsen*, one of the pioneers in neural networks development, describes *neural network*[6] as:

"...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs."

Artificial Neural Networks (ANNs) are gross simplifications of real (biological) networks of neurons inspired by the structure of the brain. Artificial neural network (see Figure 2) has each processing element (the neuron) receiving inputs from the other elements, the inputs are weighted and added, the result is then transformed (by a transfer function) into the output. The transfer function may be a step sigmoid, or hyperbolic tangent function, among others.

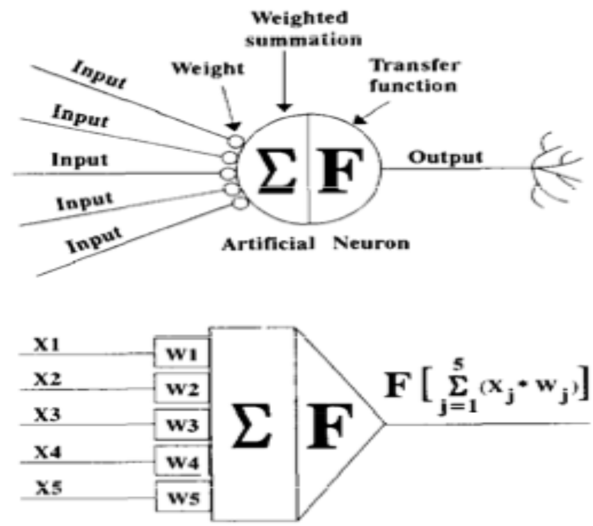

Figure 2. Functions of an artificial neuron.

Applications of artifical neural network varies from system identification and control (vehicle control, process control), game playing and decision making (chess, racing), pattern recognition (radar system, face identification, object recognition), sequence recognition (speech, gesture, and handwritten text recognition), medical diagnosis, financial applications, data mining, visualization and e-mail spam filtering.

III. THE PROCESS

Logistic regression cannot form more complex hypothesis as it is only a linear classifier, more features can be added but that can be very expensive to train. We will implement a multilayer neural network or multilayer perceptron (MLP), also known as feed-forward neural network, artificial neural network (ANN) or back prop network instead to represent complex models that form non-linear hypothesis.

A. Model Representation

An artificial neural network (ANN) is a computational model that attempts to account for the parallel nature of human brain. Network of highly inter-connecting processing

---

[5] S. Brunak and B. Lautrup, Neural Networks, Computers with Intuition, World Scientific, Singapore, 1990.

[6] Neural Network Primer: Part I by Maureen Caudill, AI Expert, Feb 1989.

elements (neurons) operates in parallel, and are truly inspired by biological nervous system. The connections between elements determine the network function and a subgroup of processing element is called a layer in the network. These layers are defined as input, output and hidden layers. Hidden layers helps stabilize the network driven by the activities of input layer.

Figure 3. Represents a typical 3-layer neural network which can be trained to perform a particular function by adjusting the values of the connections (weights) between elements.

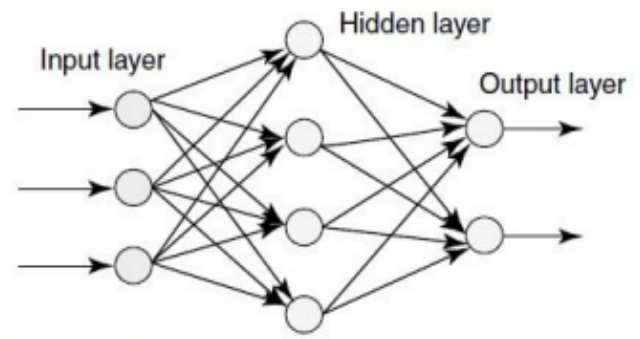

Figure 3. Multilayered Artificial Neural Network.

B. Feedforward and Cost function

Feedforward elaborates how a neural network processes and recalls patterns. The algorithim draws initial input & neural network into the test network and advances input through the network. The output is generally a vector for multilayer neural network.

For a neural network regularized cost function is as follows:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

For neural networks cost function is a generalization of the above equation, so instead of one single output, k outputs are generated.

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log(h_\Theta(x^{(i)}))_k + (1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2$$

Cost function outputs a k-dimensional vector where $h_\theta(x)$ is a k dimensional vector, such $h_\theta(x)_i$ refers to the i^{th} value in that vector. The first half infers sum for each position in output vector for every training data which is an average sum of logistic regression. Second half, is the regularization term also called a weight decay term.

To find the parameters θ which minimize J(θ) we can use one of the algorithms such as the gradient descent. For the given (Figure 4) vector implementation of forward propagation algorithm operates as follows:

$a^{(1)} = x$
$z^{(2)} = {}_{\theta}^{(1)}\, a^{(1)}$
$a^{(2)} = g(z^{(2)})$ (add $a_0^{(2)}$)
$z^{(3)} = {}_{\theta}^{(2)}\, a^{(2)}$
$a^{(3)} = g(z^{(3)})$ (add $a_0^{(3)}$)
$z^{(4)} = {}_{\theta}^{(3)}\, a^{(3)}$
Output: $a^{(4)} = h_\theta(x)\, g(z^{(4)})$

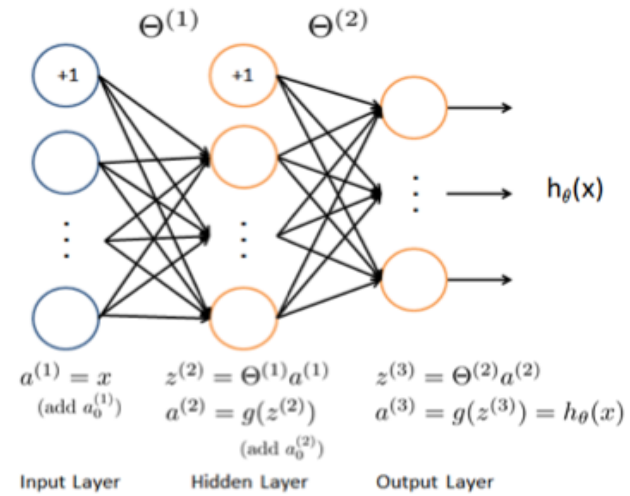

Figure 4. Feed Forward neural network model.

In general, for a feedforward operation, nonlinear multilayer networks involve input units, hidden units and output units. Hidden units enables to express more complicated nonlinear functions extending classification to greater expressive power.

C. Backpropagation

Backpropagation derives a leraning rule for the input–to-hidden weights known as credir assignment problem by computing errors for each hidden units.

Network has two modes of operation[7]:

- The feedforward operation presents a pattern to the input units and pass signals through the network in order to get outputs units (no cycles).

- The supervised learning modify network parameters (weights) to reduce distance between the computed output and desired output or target value.

Given a training example ($x^{(t)}$, $y^{(t)}$), first perform a *forward pass* to compute all activations throughout the network (including the output value of the hypothesis $h_\theta(x)$). Then for each node j in layer l, compute an *error term* $\partial_j^{(l)}$ that measures how much that node was responsible for any errors in the output. For any output node, we can directly measure the

[7] Class Notes, Week 11, CPE646-10v2

difference between the network's activation and the true target value, and use that to define $\partial_j^{(3)}$ (since 3 is the output layer). For hidden layers compute $\partial_j^{(l)}$ based on weighted average if the error terms of the nodes in layer (l+1).

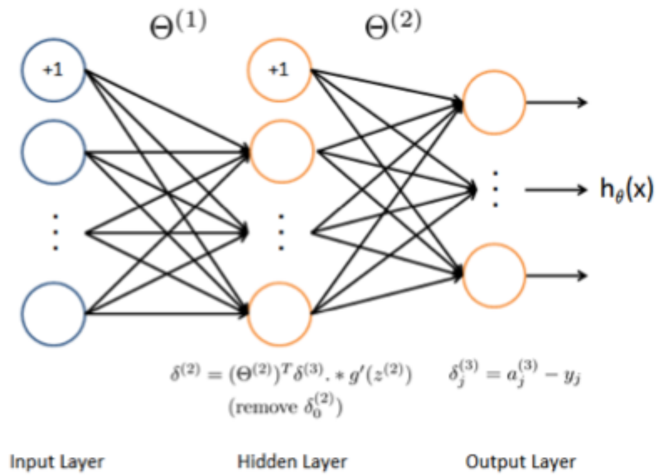

Figure 5. Backpropagation.

<u>Backpropogation Algorithim</u>:

Step 1: Set input layers values ($a^{(1)}$) to the t-th training example $x^{(t)}$. Perform a feedforward pass computing the activations ($z^{(2)}$, $a^{(2)}$, $z^{(3)}$, $a^{(3)}$) for layers 2 and 3.

Step 2: For each ouput k in layer 3 (output layer), set:

$$\partial_k^{(3)} = a_k^{(3)} - y_k;$$

where $y_k \in \{0,1\}$ indicates whether the current training example belongs to a different class ($y_k = 0$).

Step 3 : For the hidden layer l = 2, set

$$\partial^{(2)} = (\theta^{(2)})^{T(3)} .* g'(z^{(2)})$$

Step 4: Accumulate the gradient for this example, using the following formula:

$$\Delta(l) = \Delta^{(l)} + \partial^{(l+1)} (a^{(l)})^T$$

Step 5: Obtain the (un-regularized) gradient for the neural network cost function by dividing the accumulated gradients by (1/m) :

$$\partial / \partial \theta_{ij}^{(l)} = D_{ij}^{(l)} = 1/m\, \Delta_{ij}^{(l)}$$

D. Regularization

In machine learning regularization refers to the introduction of additional information to avoid over fitting (or solve ill-posed problem). This information is usually a penalty for complexity e.g. restriction for smoothness or bounds on the vector space form. To account for regularization, an additional term is added after calculating gradients using Backpropagation. Specifically, after we have computed $\Delta_{ij}^{(l)}$ using Backpropagation, we compute regularization as :

$$\partial / \partial\theta_{ij}^{(l)} .\, J(\theta) = D_{ij}^{(l)} = 1/m\, \Delta_{ij}^{(l)}$$

$$\partial / \partial\theta_{ij}^{(l)} .\, J(\theta) = D_{ij}^{(l)} = 1/m\, \Delta_{ij}^{(l)} + \lambda/m\, \Theta_{ij}^{(l)} \text{ , for } j \geq 1$$

The regularization parameter λ & number of training steps plays a pivotal role in the performance of neural network. Since neural networks are very powerful models that can form highly complex decision boundaries, without regularization, it is possible for a neural network to *over fit* a training set so that it obtains close to 100% accuracy on the training set but does not as well on new examples that it has not seen before.

IV. RESULTS

A 3-layer Neural Network has been implemented for diagnosis of Wisconsin Breast Cancer Data. In particular the following approach has been carried out:

1. Read the data file.
2. Split data into training and test set.
3. Specify Neural Network parameters (initialize random weights)
4. Minimize the cost function
5. Compute performance metrics: Prediction accuracy on training & test sets.
6. Compute Confusion Matrix.
7. Compute specificity, sensitivity & test error.

The following scatter plot visualize breast cancer dataset into Malignant & Benign.

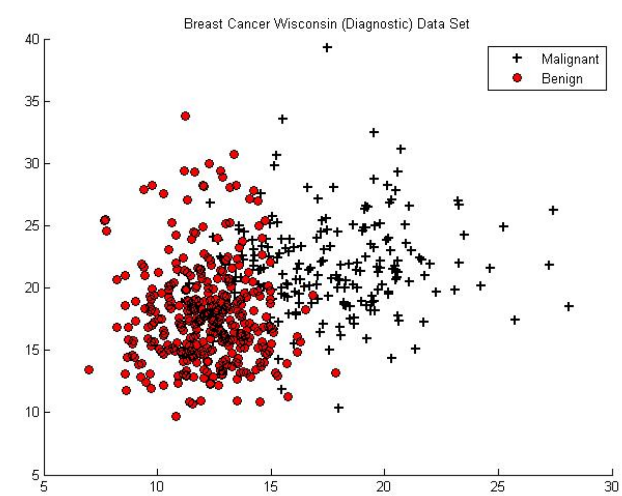

Figure 6. Wisconsin Breast Cancer Data Plot.

With a *training factor* of 0.75 data has been randomized into training & test data set.

Total Data	Training Set	Test Set
569	427	142

A 3-layer NN is applied to the above data:

Case 1:

Lambda	Hidden Layers
1	40

The training & test set accuracy are.

Iterations	Training Set Accuracy	Test Set Accuracy
500	93.911007	96.478873
450	93.676815	90.140845
400	93.911007	92.957746
300	93.911007	88.028169
200	92.740047	89.436620
100	91.803279	92.253521
50	91.803729	93.661972

Also the sensitivity, specificity, test error & total time taken for the process are:

Sensitivity	Specificity	Test error	Time (s)
0.960784	0.960784	0.3592	64.218750
0.830189	0.94382	0.3732	48.546875
0.827586	1	0.4085	39.718750
0.851852	0.897727	0.3803	30
0.8	0.945652	0.3521	21.656250
0.86	0.956522	0.3521	11.87500
0.88	0.967391	0.3521	6.421875

Case 2:

Lambda	Hidden Layers
0.5	25

The training and test set accuracy are:

Iterations	Training Set Accuracy	Test Set Accuracy
500	93.911007	94.366197
450	93.208341	93.661972
400	93.911007	88.732394
300	94.613583	93.661972
200	93.208431	90.845070
100	93.316159	90.140845
50	91.803729	88.021869

Also the sensitivity, specificity, test error & total time taken for the process are:

Sensitivity	Specificity	Test error	Time (s)
0.886364	0.969388	0.3099	60.078125
0.887097	0.975	0.4366	41.468750
0.78125	0.974359	0.4057	42.203125
0.96	0.923913	0.3521	34.562500
0.955224	0.866667	0.4718	20.406250
0.808511	0.947368	0.3310	14.687500
0.7545763	0.975904	0.4155	8

Case 3:

Lambda	Hidden Layers
1.5	30

The training and test set accuracy are:

Iterations	Training Set Accuracy	Test Set Accuracy
500	92.740047	90.845070
450	93.208431	90.845070
400	94.613583	89.436620
300	94.145199	89.436620
200	91.100703	90.845070
100	91.569087	91.549296
50	91.569087	90.845070

Also the sensitivity, specificity, test error & total time taken for the process are:

Sensitivity	Specificity	Test error	Time (s)
0.787234	0.968421	0.3310	50.265625
0.836735	0.946237	0.3451	46.734375
0.8125	0.93617	0.3380	37.25
0.819672	0.950617	0.4296	30.406250
0.811321	0.966292	0.3732	23.031250
0.839286	0.965116	0.3944	11.718750
0.8	0.923913	0.3521	6.234375

The above Tables, summarizes all the training set, test set accuracy over a number of iterations with varying lambda & hidden layers along with senstivity, specificity, test error & total time taken.

The training set and test set accuracy along with sensitivity and selectivity plots can be visualized as follows:

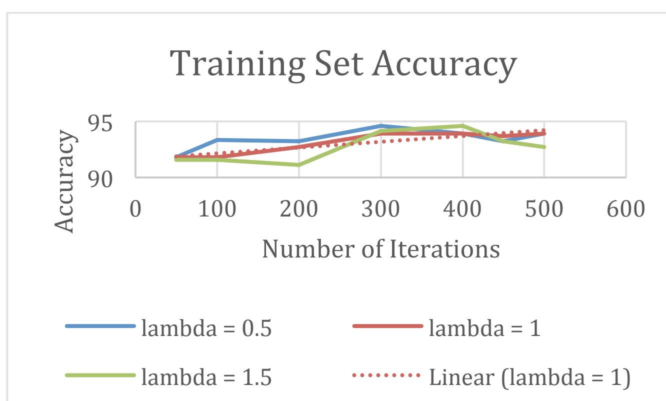

Figure 7: Training Set Accuracy.

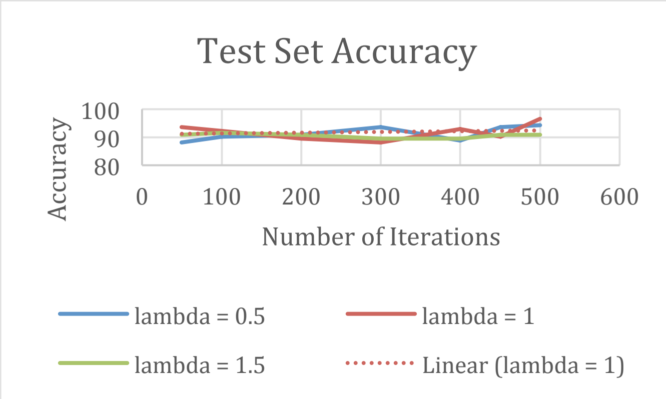

Figure 8: Test Set Accuracy.

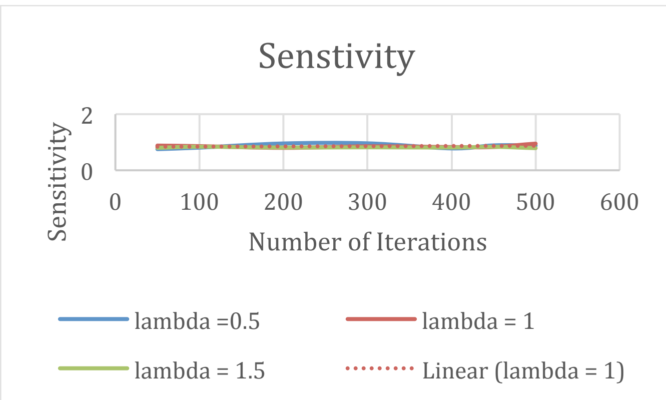

Figure 9: Sensitivity

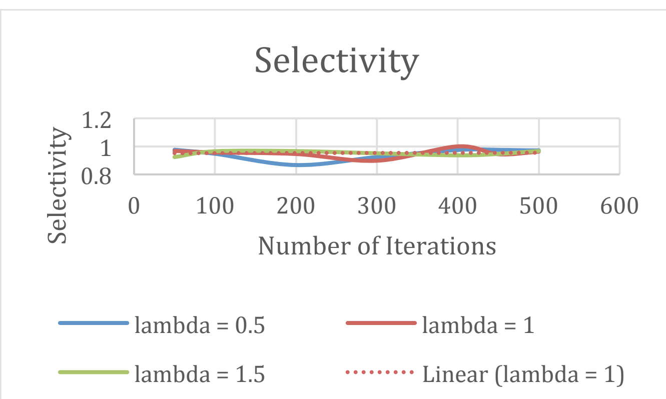

Figure 10: Selectivity

From the above data it can be inferred that training set accuracy is about 94% for varying values of regularization factor lambda and maximum iterations along with different number of hidden layers.

The designed neural network was also tested for a random data to analyze the presence of type of cancer viz. malignant & benign using the derived theta values i.e. weight matrices calculated from neural network parameters.

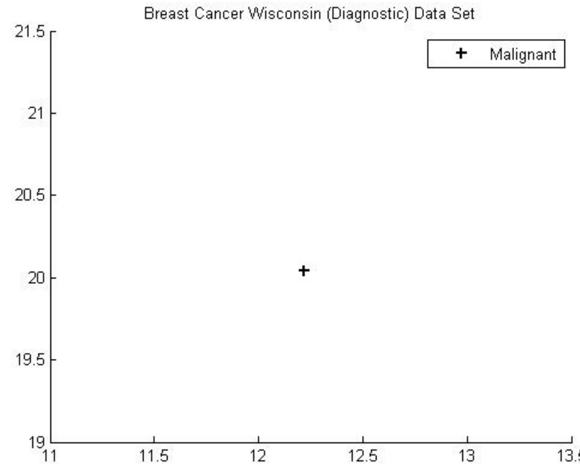

Figure 11. Malignant for a random dataset, when tested for type of cancer with the designed Neural Network.

V. Conclusion

Neural network approach to aide diagnosis of breast cancer as compared to traditional & less evasive clinical methods like surgical biopsy has high precision and accuracy. With an accuracy of training set high as 94% NNs can play an pivotal role in diagnosing malignant cancer cells. Also, training error for the large value of hidden layers[8] become small because networks have high expressive power and become tuned to the particular training set.

VI. Refrences

[1] R. C. Gonzalez, "Object Recognition," in *Digital Image processing*, 3rd ed. Pearson, August 2008, pp. 861-909.

[2] R. Dybowski and V. Gant, Clinical Applications of Artificial Neural Networks, Cambridge University Press, 2007.

[3] Parker S. L. et al, "Cancer Statistics", 1997 *CA-A Cancer Journal for Clinicians*, 47:5-27, 1997.

[4] http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

[5] S. Brunak and B. Lautrup, Neural Networks, Computers with Intuition, World Scientific, Singapore, 1990.

[6] Neural Network Primer: Part I by Maureen Caudill, AI Expert, Feb 1989

[7] Class Notes, Week 11, CPE646-10v2

[8] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", 2nd Edition, Wiley Publications, 2001. 6.8.7 - Hidden Layer Units, Chapter 6: Multilayer Neural Networks.

[9] Artificial neural networks and their business applications, Eldon Y. Li, Institute of Information Management, National Chung Cheng Unioersity, Taiwan.

[10] Artificial Neural Networks: A Tutorial, Jain, A.K. Michigan State Univeristy, MI, USA; Jianchang Mao ; Mohiuddin, K.M.

[11] "A statistical evaluation of neural computing approaches to predict recurrent events in breast cancer", Florin Gorunescu, Marina Gorunescu, Elia El-Darzi and Smaranda Gorunescu, Intelligent Systems, 2008, IS '08, 4th International IEEE Conference.

[12] Stanford Machine Learning, Prof. Andrew N.G. http://ml-class.org

---

[8] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", 2nd Edition, Wiley Publications, 2001. 6.8.7 - Hidden Layer Units, Chapter 6: Multilayer Neural Networks.