Automated Diagnosis of Breast Cancer Using Artificial Intelligence through the Implementation of Neural Networks

**Abstract**

The challenge of diagnosing any type of cancer is that the no single test can accurately succeed. Diagnostic testing is essential to successfully evaluate the health of an individual and determine whether or not the symptoms are caused by cancer or another disease. Diagnostic imaging is a useful technique to produce an internal picture of the body in order to analyze its structure. However, it is still up to the medical professional to successful analyze the images and determine whether the individual has cancer. By analyzing the data taken from the imaging with neural networks, the analysis can not only be made more efficient, but also one can minimize the error that occurs during diagnostic test.

The purpose of the project is to implement a successful neural network with back propagation to analyze a breast cancer numerical dataset. It also evaluates the efficiency or the network as it is influenced by different conditions. The efficiency is gauged by the error percentages accumulated by the network. Furthermore, statistical analysis is applied to the network in order to analyze the effectiveness of the neural network.

The project showed that…

Background Research

*Behavior exhibited by function modeling the dataset cannot be modeled by algorithmic approaches.*

Neural networks take a different approach to problem solving than conventional computers, which use an algorithmic approach where the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve.

Neural networks on the other hand learn by example and they cannot be programmed to perform a specific task. They are used to approximate an unknown function. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable. On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to solved must be known and stated in small unambiguous instructions. Neural networks and conventional algorithmic computers complement each other in order to perform at maximum efficiency.

*Potential Sources of Datasets that can be used for a neural network*

Any dataset is possible. However, having a vast dataset with many variations in it will exemplify the adaptive nature of the neural networks.

*Advantages of a Neural Network*

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organization: A neural network can create its own organization or representation of the information it receives during learning time.
3. Real Time Operation: Neural network computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage

*Artificial Neural Networks are based on a mathematical interpretation of the neural structure seen in the animal brain*

Neural networks are a form of artificial intelligence that was inspired by the structure of biological nervous systems and how they process data. Both the brain and neural networks are composed of a large number of processing elements, referred to as neurons. These neurons are interconnected by weights or axons and dendrites, which work together with the neurons to a solve problems. Both the network and any organism's nervous system learn by example. These two structures update their weights or connections in response to the input of the training data entered in and the output desired.

However, biological neurons can have up to 10,000 individual inputs, an artificial network rarely reaches the complexity of the brain. The complexity of the existing neural network is limited by the computing power of the computers or artificial systems in place today.

*Feed Forward Neural Network*

The feed forward neural network is the simplest type of artificial neural network created. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

Feed forward networks are the most popular and widely used structures to model functions that reflect the dataset. They are also known as multi-layer perceptron.

*Back-propagation*

In principle, back-propagation provides a way to train networks with any number of hidden units arranged in any number of layers.

Algorithm

By using the data, the algorithm trains a multi-layer feed forward network by gradient descent to approximate an unknown function. The overall gradient with respect to the entire training set is the sum of the gradients for each pattern.

Definitions**:**

|  |  |  |
| --- | --- | --- |
| the error signal for unit j: |  | https://lh3.googleusercontent.com/8fBTbvpb8nBGMSLnXnRhc2CAn5cP96_DrHuT6TPHtWbhceYGUndBOL5DWeY1cnB2J-sVjYcm2cGOvOcXBFqmUD8_et9GWOz3J2pya4cvqYmG3wZHljArjhBL7w\* |
| the (negative) gradient for weight wij: |  | https://lh4.googleusercontent.com/NyywYiLSuB8KwNy_gy1osBeQ7QmDjxvcQ6gRYUYUmqAZw0e1gvDMsWkN_coxcxAqQQrpw6Lfm7HGBD6aMIvcsuMkkSVC4j__oZ0I14bOpWOyfVYlfIKzXNy2og\* |
| the set of nodes anterior to unit i: |  | https://lh4.googleusercontent.com/GGDB3OHH-uaQI3N-TBCSlvYrCw9UIm8rRgfZ9plamc3HGMiBQK7At2xiAsoihw8ZGJX33V3YwG-CwGueZ4LYUfvH8TQW97sLojnhYv8XLQqXCrXAgCA85KQR8g\* |
| the set of nodes posterior to unit j: |  | https://lh5.googleusercontent.com/DAPege7Dw4Dp67fbI3R7tY2sZEvbcb_Ty6BsEpqUNWvfPZCEV7nP-oS0Ft8fdFJOWIDj_si4hsPvvGc09POcuFbhqtSjdl6t-48h_-ZojgRmHZDxUVsQZ5DyMw\* |

Expand the gradient into two factors by use of the chain rule:

https://lh4.googleusercontent.com/Zw8qEhcduA3oTVuYOnrwDLFFh9-h8b1LcJs-XlRldS-ChG28FEuptzRNGiaWWZJCFS1zXDBFHGAGRr5-4GmclDG3UJDdWVkFHaBqYfPYT_Kput02_Sv30mZAdw

https://lh6.googleusercontent.com/QFjeReNeeQfJV9rd1b9zd0DlQJqyw89F7-FVX9QpK7M0OYG_-sqX9LEWHavgr43b7qEq7d3QXOl00Win7wal3SvTHpL3vB18hXPJNeV-R_6ugrXQ6gSQva-4TQ

https://lh6.googleusercontent.com/9wEnXgSwIVT3-fVjoMJiFD_MJXpotBsamMSCHthk1xiAr0O_BtWpWFdOuulJxjOkd2WY-6NFpAHq65m9nP0mfMjn0SFD4MhzdXYaKra5Zi-vcGhoUXoVvK-cmQ.

Forward activation: The activity of the input units is determined by the network's external input. For all other units, the activity is propagated forward:

https://lh4.googleusercontent.com/wqaCymWqAgHQ3PIVbqX85UvO3hO9QK9aDi04jXRBDmxnFacL8vrDc0S9vSiEImG940fOMeZIgaG9dBjhn9RhXbebxY0qyZo51o2QJKanxXZdR4iUzryB8PUzxg

The activity of all its anterior nodes must be known.

Calculating output error. Using the sum-squared loss

https://lh5.googleusercontent.com/bnd_oY1erqQN-5kiQ54W3rABVcRUn3RMq8tSLcmdyi0XqV2X9Qp5W5KGPm4GvTJZOSeqDDfnx-Ul0M8wqsmmiXJ4yv7QjfF3erwNbf3V8KWh2K6hYw7AFoIUlQ

the error for output unit o is simply

https://lh3.googleusercontent.com/a3-q0MP2jsgRrd0BAcIMnzSM3j10R7LeKSB6krweSmOppgzSwXTQyg_njOE_b8HdbrUlIK00G63cqR_CsopJB0vsW8jiI1joN8LJWzKI_r7O3R5BHLcigYQnBw

Error back propagation. For hidden units, we must propagate the error back from the output nodes. Using the chain rule, expand the error of a hidden unit in terms of its posterior nodes:

https://lh4.googleusercontent.com/n4EhBhaDeeKDb27XosFZ3nVWHW6eS5rU4S8G94dwBsYykwiJXxjSslL8pQ3cGkKPFWznYwcF1wPG6ABotcgRPaF9KatPpBSVW5nh8EoqbB2LTQhxrl8sNIksjQ'

https://lh5.googleusercontent.com/arYo0xjoAlhe78g4vzrjKq4FDwPwJhF2sYkVzcXHtgokJ92-hF-QN6DhFG6YFBPm9z_cR-K0ZsIxFZ1jsmGN3cbdwTm2WnAyr5jZw3zbAN3RuQ5_pznumGiFkA

https://lh5.googleusercontent.com/KitXcss6ErYOQXbS67dfkFTocb9O7jklwD19q1aB_y6Q5bLVc45k6AbZBnbb479x1e0WcKFZ3FQwDjOf7IHEcVZ4rWLO-lNd7qoRzPDAXUiUSQ37IYsCmYKG8g

https://lh3.googleusercontent.com/MhZCZn5FsNxKeQqeORxq3Y5ls_aVw3pkvzUvYM4b4noURMKtRci3zeyaWamb_Ee72kcX6NTVqAFWp-P4oCZr4u7pLIOtrzFb8Ar65Bq-gXuvGnFIG6RXXzdTWw

https://lh3.googleusercontent.com/ghYVRN9Y2ubYsDTS3pfvg4z55LHUOUBYOkmZTfHoCLsYExI44tlEJWLtDoM3wgVUjyEqDEaeIIHLnknQSX6YJIrvpdwtZOaCGtdDm8jrfExQM571LlJ3VnL7pA

In order to calculate the error for unit j, we must first know the error of all its posterior nodes.  
The equation for the change in the weight relative to the error is including momentum is given by the equation.

https://lh4.googleusercontent.com/TelLlSHZpRD1bwedZ89ICRpRbj8r5xhUC8AyoPPsjxR0gpL9w19xIsZSLctHJc6ZhQPEZ_iSZHoNGVy4axsZHG9JB3E72Jj1n18Ck5Djw11YnifCQL7pqX_Glg

where mu is the momentum and t is the epoch

*Nudging*

Nudging was implemented into the network. Once the difference between ten epochs was less than .0001, then the weights randomly adjust based on the gaussian distribution. The rationale is that once the change becomes so small, then the network has become stuck in a local minimum without the ability to escape and also has not converged yet. Thus, in order to converge at the predetermined minimum error value, nuding must occur to randomly distribute the weights to try to reach the minimum.

**Inputs of the Data Points**

1. Clump Thickness                1 - 10  
2. Uniformity of Cell Size       1 - 10  
3. Uniformity of Cell Shape    1 - 10  
4. Marginal Adhesion              1 - 10  
5. Single Epithelial Cell Size   1 - 10  
6. Bare Nuclei                    1 - 10  
7. Bland Chromatin                1 - 10  
8. Normal Nucleoli                1 - 10  
9. Mitoses                        1 - 10

Clump Thickness: Benign cells tend to be group in a monolayer, while cancerous cells are often grouped in a multilayer.

Uniformity of Cell Size/Shape: Cancer cells tend to vary drastically in size and shape, thus a lower uniformity correlates with a higher possibility of cancer cells.

Marginal Adhesion: Cancer cells tend to not stick to one another as well as normal cells, so less adhesion correlates to a higher malignancy.

Single epithelial Cell Size: The size is related to uniformity, but epithelial cells enlarged may be malignant.

Bare Nuclei: It is an index of nuclei not surrounded by a cell, which is present in malignant tumors.

Bland Chromatin: Uniformity of “texture” appears in benign cells, while malignant tumors are typically coarse-textured. A lower number corresponds to more unity.

Normal Nucleoli: The rate of occurrence of normal nucleoli; abnormal nucleoli indicate possible mutated DNA, thus possible genetic expression for cancer reproduction. Thus the smaller the rate of occurrence, the larger the chance of malignancy.

The dataset used had 9 input attributes, each from a range of 1 to 10. There were a total of 699 data points. However, 16 of the points had inconsistencies where a question mark stood in place of a number. The 16 data points were thus excluded from both the training and the testing of the neural network.

While the data had the output of 2 for benign and 4 for malignant, during the testing of the network these numbers were changed to 0 and 1 for benign and malignant respectively. This is because the logistic function can only output from a range of -1 to 1. For the actual dataset, 65.5% were given as benign and the other 35.5 percent were malignant.

The source also claimed that there is also a 5% discrepancy in the dataset.

**Data processing**

The training dataset was altered first by randomly excluding 68 data points, or 10% percent, of the 683 original dataset. 10 different neural networks were created by training them on the training dataset and their weights were then saved. The weights for each neural network were then saved. These sets of weights are radically different for each other because of the existence of local minima, random weight space, and a preset convergence at .004% error. The 68 data points excluded were used as the testing data points. These data points were independently ran through the networks and then the probability of malignancy was recorded. A step function was then implemented to heavily weight the results of the network towards the malignant output. It the output was greater than .05 then the network would automatically consider the output for that data point to be malignant. Then, the output of the network was compared to the desired actual output. The average error for a single specific data point over all 10 networks next. Furthermore the average error for the total network is calculated by the average of the error for each data point. These calculations result in the total error of the network. The entire equation is given by,



**Conclusion**

The final total error calculated for the neural network is 3.2%. However, one must consider that the error in the original dataset was 5%. Thus the network was able to adapt to the error yet not perfectly. Thus while the network is a good approximation for a dataset is unable to be modeled by an algorithmic approach because of the complex patterns, it is still not a perfect representations. This randomness, caused by the random starting points of the weights centered on the zero of a Gaussian distribution is an extremely frustrating process to deal with. For the future, a heuristic way to place the starting points of the weights rather than from random values would be a more desirable way to analyze the data. The gradient method used to determine the values of the weights is not very accurate because it locates only local minimum as the threshold error value is set at .44%. Finding the global error minimum would be much more accurate. However, the problem with the global minimum is that the network might be over fitted to the data. Thus it is only trained to be able to recognize the data set it was trained on and lose its adaptive nature to recognize other potential data points. The lost of the adaptive nature of the network renders it useless as a potential way to model complex functions that are unable to be recognized by the human brain or conventional algorithms.

The step function is an extremely important part of the data processing. It is mainly used to recognize potential outliers. By weighting the output significantly, the network also becomes more accurate. Because the field the network is being applied to allow for no room for error, minimizing the amount of error received is more beneficial. It is better to have a false positive than to have a false negative because of the life-threatening issue of breast cancer. Consequently, a stepwise function must be implemented and significantly alter the output to reflect a more malignant outcome.

The 10 networks tested in the project also were very different from one another. This is caused by the random starting weight values of each network. Even though the values are centered near zero for the range, they are still randomly placed according to the Gaussian distribution. The differences reflect the self organization nature of the network and also the variance that occurs during convergence. The network also has many potential uses. It is not only applicable to breast cancer but any dataset that is or isn’t able to be modeled by conventional method. It has been used to create an app which allows for mobility and wide-range use of the network in everyday cases or by medical professionals.

<http://www.mdanderson.org/patient-and-cancer-information/cancer-information/cancer-topics/detection-and-diagnosis/diagnostic-tests/index.html>

<http://cancer.stanford.edu/information/cancerDiagnosis/>

<http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html>

<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>

<http://www.willamette.edu/~gorr/classes/cs449/backprop.html>

William H. Wolberg and O.L. Mangasarian: "Multisurface method of   
     pattern separation for medical diagnosis applied to breast cytology",   
     Proceedings of the National Academy of Sciences, U.S.A., Volume 87,   
     December 1990, pp 9193-9196.

<http://www.grappa.univ-lille3.fr/~torre/Recherche/Experiments/Datasets/#breast-cancer>