**Improving Data Analysis in Astronomy through the use of**

**Multilayer Perceptrons**

For

ICT-4410-1: Data Warehouse Design

Frederick McGovern

University of Denver University College

May 31, 2015

Faculty: Catherine Wilson

**Abstract**

Data storage capacity and CPU processing power have dramatically increased each year as predicted by Moore’s Law. As a result, each year more advanced observational equipment is being developed, leading to a significant increase in the amount of astronomical data needing to be analyzed. **Since the analysis of this data covers a vast collection of interests and importance, such as the identification of exoplanets to assisting in searching for near earth asteroids, various citizen science projects have been developed in order to increase the amount and the speed at which this data is being classified. While these endeavors are important, it is a fact that the accumulation of data is far outpacing the human capacity to analyze it. To help scientists quickly discover the new and unusual astronomical objects that exist, it is important to develop machine learning methods to reduce the amount of data required to be humanly analyzed. One possible solution is to use artificial neural networks (ANN). The multilayer perceptron is a specific artificial neural network model that is exceptionally well suited for data classification, which is a large component of astronomical data analysis,**

**Table of Contents**

[Background 1](#_Toc420749928)

[Problem Statement 5](#_Toc420749929)

[Solution 6](#_Toc420749930)

[Discussion 8](#_Toc420749931)

[Recommendations 13](#_Toc420749932)

[Conclusion 15](#_Toc420749933)

[References 16](#_Toc420749934)

# Background

For thousands of years, mankind has looked to the sky for answers to a variety of questions. Early civilizations used astronomical observations to create calendars and determine the best time to plant crops. Chinese astronomers could accurately predict when an eclipse would occur by plotting the position of the moon across the sky, documented the positions of the stars onto maps and recorded the occurrence of a supernova (Way, et al. 2012, 14). Astronomers such as Galileo, Copernicus and Newton made significant contributions to the fields of astronomy, mathematics and physics by making observations and then analyzing the data collected to validate a hypothesis. Throughout history, the data collected was freely shared to others through publications or requests in writing. The cumulating of data resulted in even more discoveries. As an example, it was data collected by Tycho Brahe that allowed Johannes Kepler to postulate his laws of planetary motion. Kepler’s laws were in fact an improvement of the initial model put forth by Nicholaus Copernicus (Glymour 2012, 12-13).

The invention of the telescope dramatically increased the amount of astronomical data being collected and classified. In 1781, Charles Messier published a catalog of 103 unusual objects, which today would be recognized as galaxies, nebulae and star clusters, which he observed while looking for comets (Feigelson 2012, 3). During the early 18th century, English astronomer John Flamsteed created a catalog documenting the position of approximately 3000 stars. By the late 19th century, spectroscopic analysis of stars via telescope led to the creation of a catalog that contained over 300,000 stars. The Palomar Sky Survey, completed in 1958, led to the creation of even more catalogs with every increasing content (Feigelson and Babu 2012).

The advent of computers and digital photography once again caused a giant leap forward in the amount of data being collected. It also had a profound impact on the way data was being shared. Instead of requesting data via correspondence or actually traveling to an observatory, an astronomer could, from the comfort of his office, access datasets from various institutions and download them to his personal workstation for data analysis. As computational processing power increased and costs for CCD detector systems and other hardware decreased, additional and larger telescopes were built and surveys conducted. Satellites designed to study all wavelengths from infrared to x-ray transmissions were commissioned. Astronomers now had a plethora of data, in many cases taken at various times or collected at various wavelengths, to make comparisons against. Instead of data on millions of stars, the Sloan Digital Sky Survey (SDSS), which began operation in 2000 and took repeated surveys of approximately 35 percent of the night sky, collected information such as position and brightness on a billion stars and galaxies. This survey, produced 200GB of data each night and the resulting database was in excess of 50TB as of 2012 (Feigelson and Babu 2012). The increase in data had also begun to effect the performance at the data centers that store the data. The NASA Infrared Processing and Analysis Center, which stores the data for all infrared missions, in recent years began digital curation of all data from the Spitzer Space Telescope and the Wide-Field Infrared Survey Explorer (WISE). As can be seen from the accompanying graph (Figure 1), the combined data sets for the Spitzer and WISE mission dramatically exceed the volume of all previously archived mission and projects, of which there were 35 in total (Berriman and Groom 2011).

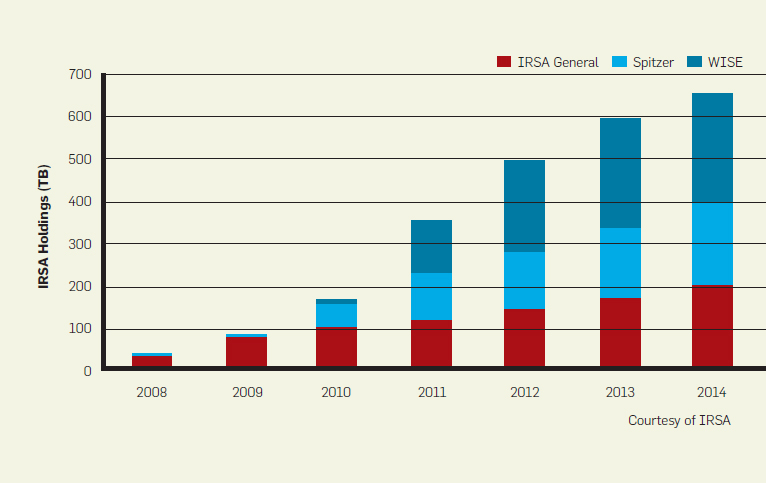


Figure 1

The dramatic increase in data led astronomers to the conclusion that there was simply too much data for them to single-handedly analyze. This led to the idea of outsourcing specific data analysis tasks, such as galaxy classification or locating exoplanets to the general public. One of the first attempts in the development of public interactive websites for public data analysis was the Galaxy Zoo website designed by Phil Murray and Dan Andreescu. The website, which utilized data from the SDSS, launched in July 2007. Visitors to the site were asked to classify a galaxy as either “Elliptical’ or “Spiral” and then to determine if the spin of the galaxy appeared to be clockwise or counter clockwise. This site had a tremendous response from the community. Within 3 hours of the site launch, classifications where exceeding the initial capability of the data servers. After additional resources where brought online, classifications increased to 20,000 per hour after 12 hours of site launch and rose to 60,000 per hour after 40 hours of launch. By April 2008, over 100, 000 volunteers had classified the 900,000 galaxy images provided by the SDSS an average of 38 times (Forston, et al. 2011, 217-219).

The success of Galaxy Zoo led to the creation of the Zooniverse, a cloud based web portal that was developed to act as a gateway to other site looking for assistance from the public in data analysis. From here, an interested volunteer can jump to the PlanetHunters website to search for exoplanets using data from NASA’s Kepler mission or they can visit SolarStormWatch to help astronomers spot and track solar eruptions (Space 2015).

As more and more data was collected, astronomers also realized that a new method for accessing and analyzing the data would need to be found. This led to the creation of the Virtual Observatory (VO). The VO provides web based tools to allow astronomers to search and conduct analysis of archived data across the globe. International collaboration amongst the data providers has led to the creation of shared standards and protocols ranging from data access to data modeling. Query of the geographically distributed data sets is accomplished via an online VO service called SkyQuery. Additional VO tools include applications utilizing machine learning to assist in data classification and detection of unusual objects (Borne 2012, 450-452).

While these efforts are exemplary in their attempt to provide some level of analysis of the astronomical data that is being produced daily and both should be included as part of the overall solution to the problem, the fact remains that another method of data analysis must be found. The Space Telescope Science Institute reports that there are more research papers published that utilize archival data than those that utilize newly discovered data (Berriman and Groom 2011). Additionally, there are many more projects that will exponentially increase the amount of data being produced. For example, the SkyMapper telescope will produce 500TB of data over the course of its lifetime (Tyson and Borne 2012). The Large Synoptic Survey Telescope (LSST) on the other hand will produce 30 TB of data per night for an expected lifetime output of 60PB for raw data and 30PB for cataloged data (LSST and Technology Innovation 2015). That is just a precursor for the Square Kilometer Array which begins construction in 2018. Once completed in 2023, it is expected to produce 1PB of data every day (Budavari, Dobos, and Szalay, 2013, 19).

## **Problem Statement**

As predicted by Moore’s Law, data storage and CPU processing power have exponentially increased over the years. These advances have led to development of new telescopes and other instrumentation which has resulted in a dramatic increase in the amount of data being obtained from astronomical observation. **Research has already been hampered as queries against large volumes of data are taking too long or the data set being downloading to local researches computers is just too large.** As pointed out by Bruce Berriman and Steven Groom, over 1PB of astronomical data is publicly available for download. As such, “Without intervention, the current data-access and computing model used in astronomy, in which data downloaded from archives is analyzed on local machines will break down rapidly (Berriman and Groom 2011). A new approach to data analysis must be developed. This solution needs to dramatically reduce the amount of data needing to be humanly analyzed. At the same time, any solution developed needs to produce a high degree of accuracy when compared to the results that would be obtained from traditional human data analysis.

# Solution

Astronomers perform a wide range of tasks, many which are esoteric in nature. As a result, some astronomers cannot rely solely on expert analysis of data or the contributions made through citizen science. To handle the increases in data, machine language algorithms have started to appear in astronomical data analysis. Examples of this include support vector machines to determine photomic and spectroscopic redshifts, Bayesian analysis to determine how stars form in young galaxies and decision trees to classify galaxies and spatial data streams (Franco-Arcega et al 2013). It is clear that the incorporation of a machine learning algorithm is necessary to remove the necessity of human based data analysis but a wide variety of potential solutions exist. As the majority of astronomical data analysis is designed to characterize the data, primarily image data, the best choices would seem to be decision trees, neural networks such as the multilayer perceptron or the utilization of support vector machines.

Decision trees are certainly the easiest algorithm to understand as it is composed of a series of if-then statement. However, due to the need to introduce dimensionality in the data by comparing images taken at various times and at various wavelength, this structure would result in the need for extremely large and complex trees which requires the need for experts in the area of algorithm design. Additionally, each data point is looked at sequentially, so it is can produce spurious relationships (Nayab 2011).

A very strong argument can be made for the use of support vector machines (SVM), as numerous papers that compare the accuracy of SVM to neural networks show that SVM have a higher accuracy prediction rate. The main reason for this is that neural networks have a higher propensity to over fit. SVM utilize two adjustable parameters – the kernel width and a regulation of the classification error which helps to avoid over fitting (Ball and Brunner 2009). However, these results are highly dependent on what kernel function is chosen as different applications require different kernel functions to get reliable classification results (Zanity 2012). Other disadvantages of SVM are that due to its highly complex algorithm and memory requirements the SVM takes longer to train and test (Olson 2008). This isn’t a function of the number of adjustable parameters, SVM have fewer than neural networks, and it is just that the process of optimization of the parameters of an SVM take longer (Ball and Brunner 2009).

Astronomical data classification is usually feature based, that is we make classifications based on the features on metrics in the underlying data. The underlying data may have multiple dimensions and need to be sorted into multiple classification. Unfortunately, the SVM was designed for two class classification. That is, a SVM will find the hyper plane that produces the largest margin between the two classes. A neural network doesn’t have this restriction so while a SVM might be a better approach for a specific application, in general, due to the wide dimensionality and number of features in the data, a neural network can be used in more types of data classification. Neural networks scale exceptionally well. The neural network has the capability to add more hidden layers to the network as the complexity of the problem increases. Additionally, concerns in regards to over fitting can be addressed through the process of cross validation (Ball and Brunner 2009). Neural networks provide a good approximation for nonlinear function and are easily modified for parallel processing. Finally, neural networks do not need input by domain experts and can actually find relationships between the features that were previously unknown. (Marakas 2003).

# Discussion

To fully understand why a multilayer preceptor is a good choice for astronomical data classification, the reader should have at least have an understanding of how they work along with some historical background. The concept of an artificial neural network can be traced back to the work done by Warren McCullum and Walter Pitts at the University of Chicago. Pitts was familiar with the works of Gottfried Liebniz and combined with the ideas of the universal computing engine put forth by Alan Turing wondered whether or not the human nervous system could be characterized as a universal computing machine. Their published paper, *A Logical Calculus of Ideas Immanent in Nervous Activity*, presented a model of a neuron with had two inputs and produced a singular binary output. The output remained zero until a specified threshold value was reached via a sum of the inputs. These inputs were considered a positive integer if an excitatory signal was received by the neuron. This is a simplified version of what today is known as a logical threshold gate (Figure 2) (2.3.1 The McCulloch-Pitts Model of Neuron 2015).



http://wwwold.ece.utep.edu/research/webfuzzy/docs/kk-thesis/kk-thesis-html/img18.gif

Figure 2

Utilizing the figure above, if we had two inputs, A and B, and the threshold value required to produce an output C is two, then both inputs must receive an excitatory signal or else the output will remain at zero.. This can be represent by a simple AND logic gate (Figure 3) where the number one indicates an excitatory signal has been received by the input (Beeman 2001).

|  |  |  |
| --- | --- | --- |
| A | B | C |
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

Figure 3

This model assigned identical weights to each of the inputs. These weights were designed to apply a level of importance to the inputs and were in the range of 0< x < 1. This means that the output will be a value of 1 if the following mathematical expression is true: (I1\*W1) + (I2\*W2) + (I3\*W3) + ……(In\*Wn) >= T. Additionally, the McCulloch-Pitts neuron model also introduced the concept of an inhibitory input. If this input was present then the output would be zero regardless of whether or not the threshold value had been reached by the sum of the input values (McCulloch Pitts Neurons 2015).

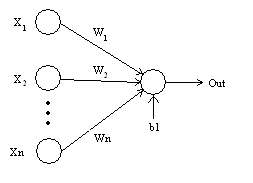
Even though the McCulloch-Pitts neuron model could be defined by a mathematical expression, it was very simplistic and limited to only allowing changes to the input weights and threshold values. It acted as digital logic circuit having the ability to implement Boolean logic function such as AND, OR, NAND and NOR but it couldn’t compute the exclusive OR function. Additionally, this model did not have the capability to learn. To get the desired output weights and thresholds had to be continuously manually changed (Beeman 2001).

In 1958, psychologist Frank Rosenblatt took the McCulloch Pitts model of a neuron and combined it with a concept of adjustable weights that had been proposed by Donald Hebb in 1949. The result was the Perceptron. The McCulloch Pitts neuron had all the excitatory inputs carry the same weight. However, Rosenblatt realized that didn’t need to be the case. It is possible that one input signal should have a stronger influence than other input signals when determining if the threshold value has been reached. Additionally, instead of only the possibility of a positive weight, Rosenblatt allowed for both negative and positive weights represented by a decimal number in the range of (-1, 1). The perceptron also adds in the concept of a bias (Figure 4) and changed the output from (0, 1) to (-1, 1). The bias is an indicator of how likely the output should be 1 so in affect is a way to adjust to the threshold value of the neuron. Mathematically this can be represented by

n

∑ XiWi + b1 = Sum

i=1



http://matlabgeeks.com/wp-content/uploads/2011/05/Perceptron.bmp

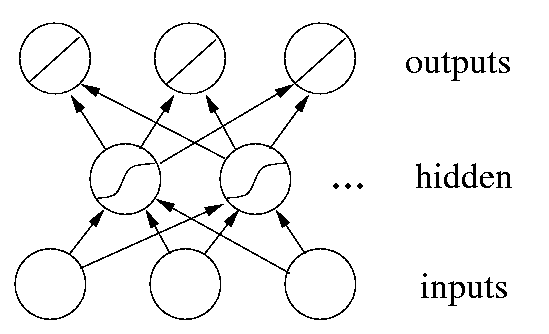
Figure 4

Mathematically we can show that the above equation, when set equal to zero, produces a hyperplane (Alpaydin 2014). We can define two classes of objects, those being C1 and C2, corresponding to output values of -1 and 1. Thus the hyperplane produced serves as the barrier that separates the two classes. By modifying the perceptron bias, you in effect shift the hyperplane (Haykin 2009).

The importance of the preceptor is that it introduced a learning algorithm. Initially, Rosenblatt would manually adjust the various weights to try to “train” the preceptor to recognize specific input signals should belong to specific classes. In 1960, Bernard Widrow and Tim Huff created the Least Mean Squares algorithm. This algorithm determines the necessary coefficients so as to produce the least mean squares between the actual output produced and the desired output (error signal) using a process known as the gradient descent method. Applying this to the preceptor allowed for an automatic correcting to the input weights based on the existing error signal. If the data provided as input is linearly separable, that is it belongs to a singular class such as found in a two input perceptron, eventually convergence will occur onto a consistent set of weights. This is known as the Perceptron Convergence Theorem. If it is not linearly separable, such as the three input perceptron, then there will always be at least one misclassification resulting in an infinite cycle of weight modifications (Haykin 2009).

While the perceptron made dramatic improvements to the original McCulloch Pitts neuron and had created a lot of interest among researchers, it still had its limitations. Chief among them was the fact that a perceptron could only solve linearly separable problems and could not solve the XOR functions. This led Marvin Minsky, in 1969, to proclaim in his paper *Perceptrons: An Introduction to Computational Geometry*, that any continuing work with the perceptron would fail due to these limitations. This led to halt in future research (Wallis 2015).

Research in self-organizing neural networks remained slow, however, small advances in the development and training of artificial neural nets continued. Paul Werbos, in 1974, introduced the mathematical foundation of a new learning algorithm (Levin 1996). This algorithm would eventually be dubbed back propagation by David Rumelhart, Geoffrey Hilton and Ronald Williams. In their paper, *Learning representations by back-propagating errors,* the authors describe a multilayer network (Figure 5) consisting of a layer of inputs units at the bottom, any number of hidden units in the middle and a layer of output units at the top.



http://www.willamette.edu/~gorr/classes/cs449/figs/mlp.gif

Figure 5

The mathematics of the backpropagation are complex but is similar to the Least Mean Square Algorithm as it calculates the sum of the squared error between the expected and actual output. A determination is then made how much the individual units have to be adjusted in weight. Changes are made first at the output layer and then in each hidden layer progressing downward toward the input layer. This process continues in an effort to get closer to a correct output (Rumelhart et al 1986). The process will continue until a certain number of complete passes through the network are reached or the percentage of errors is below a preset threshold. Current backpropagation methods have incorporated the concept of a learning rate. The learning rate is a constant that adjust the degree of change that is made to the weights. Ideally, changes should be small. If weight changes are too large, then the optimized weights could be difficult to obtain as weights continue to be adjusted either higher or lower than the optimal weight. Conversely, if the degree of change is too small, then the time it takes to reach optimal weights can become excessive. Backpropagation is a form of supervised learning primarily used in feed forward neural networks (Marakas 2003).

As previously discussed a perceptron cannot solve non-linear problems. However, when perceptrons are combined in a feedforward network that contains hidden layers between the input and output layers, can implement nonlinear discriminants when used for classification. The XOR function can be implemented using this approach. This is accomplished through the use of sigmoidal function. This provides a differential version of thresholding and is needed due to the existing learning algorithm, such as backpropagation, use the gradient descent method. Backpropagation is still utilized to train the multilayer perceptron, however, instead of the output being a linear function of the input, it is a nonlinear function (Alpaydin 2014).

# Recommendations

Obviously, the inner workings of a multilayer perceptron are complicated. When used in classification, there is some faith needed that the resultant output is the right one. However, they are “universal approximators”, that is they can approximate any function with a relatively high degree of accuracy (Oza 2012). No method of classification is perfect. While humans are very good at pattern recognition, we sometimes combine different parts of an image to create something familiar such as a face. It can be very difficult to visually tell if an astronomical image represent an actual object or if the ambient noise in the image is making us believe an unusual object exists (Scargle 2012).

If provided with a large and diverse training set, a multilayer perceptron eliminates the need for a domain expert for further data classification. Fortunately, we have established training and test sets that we can utilize thanks to the contribution of citizens to crowdsourcing projects like GalaxyZoo. Contributions made by these organizations will remain important. Due to the newer telescopes, such as the LSST, we will be gaining a new dimensionality in the data, that being time. Fluctuations can be expected to occur in the data, not only because the universe isn’t a static entity but also due to weather conditions and equipment defects. Data is constantly changing and as a result, new data will need to be introduced into the training sets to further improve overall accuracy of classification (Ball 2009).

Machine learning application to astronomical studies remains low. An analysis of the four journals that publish the majority of astronomical research showed that of the 70,000 articles published between 2000 and 2010, only 624 articles referred data mining or machine languages. However, neural networks had a high percentage of articles compared to other types of machine learning (Feigelson 2012). In fact multilayer perceptrons have successfully been utilized to determine galaxy morphology. The Hubble sequence of galaxy classification has been extended to what is called the T system. Papers published by astronomer Ofer Lahav, demonstrated that image classification of galaxies in low red shift to the T system, when classified through an artificial neural network had a degree of accuracy equal to that of the classifications done by domain experts (Ball 2009). Additional examples of the use of multilayer perceptron include star classification, cosmic ray detection, solar flare detection and supernova classification (Bloom and Richards 2012).

# Conclusion

Astronomers are overwhelmed with data. New star surveys are capturing data on stars, galaxies, exoplanets and local transient objects. The combined number of potential data points is in the billions. Human classification has of astronomical data has served us well in the past, but with the sheer volume of data that is now available, this is no longer an option, even when we add the contribution of crowdsourcing sires such as Galaxy Zoo. Additionally, astronomers need to quickly identify the unusual objects as these can lead to new discovers or even help to avoid potential impacts from astronomers. The only plausible solution to this issue is to reassign the task of classifications to machines. There are a variety of different machine learning algorithms to choose from, however, the use of a multilayer perceptron provides numerous benefits when used for classification compared to other methods. This approach uses a supervised learning method, such as backpropagation, to improve the overall accuracy of the predicted classification over time. Large training sets, help to improve accuracy and those do exists in the community.

Finally, the use of neural networks, such as a multilayer perceptron, to classify astronomical data are not new concepts. While the overall usage of machine learning algorithms in astronomy is low, the use of neural networks is at the top of what is currently successfully being utilized. It is true that machine learning algorithms in general and multilayer perceptrons specifically can be complicated to understand. Fortunately, the algorithms are based on sound mathematical methods and have a high level of accuracy. This means that an astronomer can use these methods without an understanding how underlying algorithm actually works and have confidence that the provided classification is correct.

# References

Alpaydin, Ethem. *Introduction to Machine Learning*. 3rd ed. Cambridge, Mass.: MIT Press, 2014.

Ball, Nicholas M., and Robert J. Brunner. "Data Mining And Machine Learning In Astronomy."*International Journal of Modern Physics D*, 2009, 1049-106.

Beeman, Dave. "Some Specific Models of Artificial Neural Nets." McCullogh-Pitts and Perceptron Models. October 30, 2001. Accessed May 25, 2015.

Berriman, G. Bruce, and Steven L. Groom. "How Will Astronomy Archives Survive the Data Tsunami?" *Communications of the ACM* 54.12 (2011): 52.

Borne, Kirk D. 2012 “Virtual Observatory and Distributed Data Mining” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 447-462. Boca Raton, FL: CRC Press, 2012.

Budavari, Tamas, Laszlo Dobos, and Alexander S. Szalay. "SkyQuery: Federating Astronomy Archives." *Computing in Science & Engineering* 15, no. 2 (2013): 12-20.

Feigelson, Eric D., and G. Jogesh Babu. "Big Data in Astronomy." *Significance* 9, no. 4 (2012): 22-25.

Feigelson, Eric. 2012 “Classification in Astronomy: Past and Present.” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 3-10. Boca Raton, FL: CRC Press, 2012.

Fortson et al. 2012 “Galaxy Zoo: Morphological Classification and Citizen Science.” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 213-236. Boca Raton, FL: CRC Press, 2012.

Franco-Arcega, A., L.G. Flores-Flores, Ruslan F. Gabbasov. “Application of decision trees for classifying astronomical objects.” *2013 12th Mexican International Conference on Artificial Intelligence,* 2013.

Glymour, Clark. 2012 “Searching the Heavens: Astronomy, Computation, Statistics, Data Mining, and Philosophy” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 11-27. Boca Raton, FL: CRC Press, 2012.

Haykin, Simon. "Single Layer Perceptrons - Least-Mean-Square Algorithm Perceptron." 2009. Accessed May 25, 2015. http://www.fatih.edu.tr/~aliadam/EEE544A/EEE544LMSPerceptron.pdf.

Leven, Sam. "The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting." Neural Networks 9, no. 3 (1996): 543-44. Accessed May 26, 2015. doi:0893-6080(94)00090-5.

"LSST and Technology Innovation." Large Synoptic Survey Telescope. Accessed May 18, 2015. http://www.lsst.org/lsst/about/technology.

Marakas, George M. "Machines That Can Learn." In Modern Data Warehousing, Mining, and Visualization: Core Concepts. Upper Saddle River, NJ: Prentice Hall, 2003.

"McCulloch Pitts Neurons." - The Mind Project. Accessed May 25, 2015. http://www.mind.ilstu.edu/curriculum/mcp\_neurons/mcp\_neuron\_1.php?modGUI=212&compGUI=1749&itemGUI=3018

Nayab, N. "A Review of Decision Tree Disadvantages." Brighthub Project Management. February 9, 2011. Accessed May 23, 2015. http://www.brighthubpm.com/project-planning/106005-disadvantages-to-using-decision-trees/

Olson, David Louis, and Dursun Delen. "Support Vector Machines." In Advanced Data Mining Techniques. Berlin: Springer, 2008.; 112-122

Oza, Nikunj. 2012 “Classification.” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 505-522. Boca Raton, FL: CRC Press, 2012.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-propagating Errors." Nature, 1986, 533-36. Accessed May 26, 2015. http://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop\_old.pdf.

Scargle, Jeffrey D. 2012 “Probability and Statistics in Astronomical Machine Learning and Data Mining.” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 27-36. Boca Raton, FL: CRC Press, 2012.

“Space.” Zooniverse. Accessed May 14, 2015. https://www.zooniverse.org/projects.

Tyson, J. Anthony, and Kirk D. Borne 2012 “Future Sky Surveys.” In *Advances in Machine Learning and Data Mining for Astronomy*, Eds.: Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, Ashok N. Srivastava, 595-616. Boca Raton, FL: CRC Press, 2012.

Wallis, C. "History of the Perceptron." History of the Perceptron. Accessed May 26, 2015.

Zanaty, E.a. "Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in Data Classification." Egyptian Informatics Journal, 2012, 177-83.

"2.3.1 The McCulloch-Pitts Model of Neuron." 2.3.1 The McCulloch-Pitts Model of Neuron. Accessed May 25, 2015. http://wwwold.ece.utep.edu/research/webfuzzy/docs/kk-thesis/kk-thesis-html/node12.html