Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure

Olabode Olatubosun2
Computer Science Dept,
The Federal University of Technology, Akure

Fasoranbaku Olusoga
Statistics Department
The Federal University of Technology, Akure

Fagbuwagun Abayomi
Computer Science Department
The Federal University of Technology, Oye-Ekiti

Abstract

This study presents Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure. The Maximum covariance method is divided into three phases. A large number of candidate's hidden units is considered by initializing their various weights with random values. Then the desired number of hidden units is selected amongst the candidates by using the maximum covariance. The weights feeding the output units are calculated with linear regression method. After the maximum covariance initialization, the network is trained with the resilient back propagation which is an adaptive training algorithm. The activation function in the hidden units is hyperbolic tangent function. Ten baseline variables includes, age, sex, body mass index, average blood pressure and six blood serum measurements, were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline was used. The learning machine was trained, validated and tested. The result shows the algorithm is efficient in the diagnosis of who is a diabetic patient.

Keywords: Maximum Covariance, Back Propagation, Diabetes, Hyperbolic and Diagnosis

---

IJCIR Reference Format:

Olabode Olatubosun, Fasoranbaku Olusoga and Fagbuwagun Abayomi. Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure
International Journal of Computing and ICT Research, Vol. 8, Issue 2 pp 10-21.
http://ijcir.mak.ac.ug/volume8-issue2/article2.pdf

2Author's Address: Olabode Olatubosun, Computer Science Dept, The Federal University of Technology, Akure.

International Journal of Computing and ICT Research, ISSN 1818-1139 (Print), ISSN 1996-1065 (Online), Vol. 8, Issue 2, pp 10-21, December 2014

1.0 Introduction

Perceptron is the most used artificial neuron in neural network configurations as opined in Rosenblatt [1958]. This is based on the nonlinear model as proposed in McCulloch and Pitts [1943]. In the model Rossana, Helton and Rafael [2011], neurons are signal processing units composed by a set of input connections of weights, an adder for summing the input signals, weighted by the respective synapses of a neuron, constituting a linear combiner and an activation function, that can be linear or nonlinear. The input signals are defined as $x_i, i = 0, 1, \ldots, N_i$, whose result corresponds to the level of internal activity of a neuron net_j, as defined in Eq.1, where $x_0 = +1$ is the polarization potential (or bias) of the neurons. The output signal y_j is the activation function response $\varphi(\cdot)$ to the activation potential net_j, [Silva et al. 2010].

$$net_j = \sum_{i=0}^{N_i} w_{ji} \cdot x_i \qquad \ldots 1$$

$$y_j = \varphi(net_j)$$

For a feedforward neural network (FNN), the artificial neurons are set into layers. Each neuron of a layer is connected to those of the previous layer. Signal propagation occurs from input to output layers, passing through the hidden layers of the FNN. Hidden neurons represent the input characteristics, while output neurons generate the neural network responses [Haykin 1999].

Diabetes disease diagnosis via proper interpretation of the Diabetes data is an important classification problem [Davar et. al 2012]. In this study, an attempt is made to design a framework of Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Neural Network Procedure

There are many factors to analyze to diagnose the diabetes of a patient, and this makes the physician's job difficult. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. But, this is not easy considering the number of factors she has to evaluate [Davar 2012; Polat et al. 2008]. To help the experts and helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, classification systems provide medical data to be examined in shorter time and more detailed.

2.0 Diabetes Mellitus and Diagnosis

Diabetes occurs when a body is unable to produce or respond properly to insulin which is needed to regulate glucose [Alberti and Zimmet 1998]. Diabetes is not only a contributing factor to heart disease, but also increases the risks of developing Kidney disease, Blindness, Nerve damage, and blood vessel damage. Statistics has shown that more than 80 percent of people with Diabetes die from some form of heart or blood vessel diseases. Currently there is no cure for Diabetes; however, it can be controlled by injecting insulin, changing eating habits, and doing physical exercises [Polat and Günes 2007].

Diabetes is diagnosed by an excessively high concentration of glucose in the blood occurring spontaneously or following an oral glucose challenge [National Diabetes Data Group 1979]. Most people with diabetes can be classified into one of two major types. Insulin-dependent (type I) diabetes is characterized by dependence on exogenous insulin to prevent ketoacidosis and death, by the presence of antibodies to pancreatic islet cells, and often by an abrupt onset of symptoms. Non-insulin-dependent (type II) diabetes is characterized by ketosis resistance, lack of islet-cell antibodies, and frequently an insidious or asymptomatic onset [William, David, Peter, and Robert 1983].

Some diabetic patients will not get any warning sign or symptoms. The only way to be sure is to have blood test for glucose [Irvine, Toft, Holton, Prescott, Clarke, and Duncan 1977]. The diabetic's diagnosis tests include-

a. ***Fasting Plasma Glucose:*** The fasting plasma glucose (FPG) test is the standard test for diabetes. It is a simple blood test taken after 8 hours of fasting. Results indicate:

• FPG levels are considered normal up to 100 mg/dL

• Levels between 100 and 125 mg/dL are referred to as impaired fasting glucose or pre-diabetes.

• Diabetes is diagnosed when FPG levels are 126 mg/dL or higher on two or more tests on different days.

b. ***Postprandial blood glucose test (PPB):*** This test is followed by Fasting plasma glucose test. Take good amount of food after FPG, wait 2 hours, and do the blood test again.

- Postprandial glucose level should be under 140 mg/dL.
- The value between 140 and 199mg/dL indicates pre-diabetes.
- 200 and above value may indicate diabetes.

c. ***Random blood glucose test****:* A random blood glucose test can be used to diagnose diabetes patient, in other words,

- A blood glucose level of 200 mg/dl or higher indicates diabetes

The Oral glucose tolerance test:

The oral glucose tolerance test is used for the diagnosis of type 2 diabetes. It is also used for diagnosing gestational diabetes and in conditions of pre-diabetes. With an oral glucose tolerance test, the person fasts overnight (at least eight but not more than 16 hours). Then first, the fasting plasma glucose is tested. After this test, the person receives 75 grams of glucose (100 grams for pregnant women). Blood samples are taken at specific intervals to measure the blood glucose over a period of three hours. In a patient without diabetes, the glucose levels rise and then fall quickly. In someone with diabetes, glucose levels rise higher than normal and fail to quickly come down as fast. Patient with glucose levels between normal and diabetic have impaired glucose tolerance (IGT). People with impaired glucose tolerance do not have diabetes, but are at high risk for progressing to diabetes. Glucose tolerance tests may lead to one of the following diagnoses:

• Normal response**:** A person is said to have a normal response when the 2-hour glucose level is less than 140 mg/dl, and all values between 0 and 2 hours are less than 200 mg/dl.

• Impaired glucose tolerance**:** A person is said to have impaired glucose tolerance when the fasting plasma glucose is less than 126 mg/dl and the 2-hour glucose level is between 140 and 199 mg/dl.

• Diabetes**:** A person has diabetes when two diagnostic tests done on different days show that the blood glucose level is high.

• Gestational diabetes**:** A woman has gestational diabetes when she has any two of the following: a 100g OGTT, a fasting plasma glucose of more than 95 mg/dl, a 1-hour glucose level of more than 180 mg/dl, a 2-hour glucose level of more than 155 mg/dl, or a 3-hour glucose level of more than 140 mg/dl.

3.0 Resilience Back Propagation Neural Networks.

The Back Propagation algorithm begin with computing the output layer, which is the only one where desired outputs are available, where the outputs of the intermediate layers are unavailable as presented in Graupe [2007] as follows:

Let ε denote the error-energy at the output layer, where:

$$\varepsilon \triangleq \frac{1}{2}\sum_k (d_k - y_k)^2 = \frac{1}{2}\sum_k e_k^2 \qquad \dots 2$$

$k = 1 \dots N;\ N$ being the number of neurons in the output layer. Consequently, a gradient of ε is considered, where:

$$\nabla \varepsilon_k = \frac{\partial \varepsilon}{\partial w_{kj}} \qquad \ldots 3$$

by steepest descent (gradient) procedure, a weights vector settings after iteration is given by:

$$w_{kj}(m+1) = w_{kj}(m) + \Delta w_{kj}(m) \qquad \ldots 4$$

j denoting the jth input to the kth neuron of the output layer, where, again by the steepest descent procedure:

$$\Delta w_{kj} = -\eta \frac{\partial \varepsilon}{\partial w_{kj}} \qquad \ldots 5$$

The minus (-) sign in Eq.5 indicates a down-hill direction towards a minimum. Note from the perceptron's definition that the k's perceptron's node output z_k is given by

$$z_k = \sum_j w_{kj} x_j \qquad \ldots 6$$

x_j being the jth input to that neuron, and noting that the perceptron's output y_k is:

$$y_k = F_N(z_K) \qquad \ldots 7$$

F being a nonlinear function. Substitute for

$$\frac{\partial \varepsilon}{\partial w_{kj}} = \frac{\partial \varepsilon}{\partial z_k} \frac{\partial z_k}{\partial w_{kj}} \qquad \ldots 8$$

and, by Eq.6:

$$\frac{\partial z_k}{\partial w_{kj}} = x_j(p) = y_j(p-1) \qquad \ldots 9$$

p denoting the output layer, such that Eq.8 becomes:

$$\frac{\partial \varepsilon}{\partial w_{kj}} = \frac{\partial \varepsilon}{\partial z_k} x_j(p) = \frac{\partial \varepsilon}{\partial z_r} y_j(p-1) \qquad \ldots 10$$

Defining:

$$\Phi_k(p) = -\frac{\partial \varepsilon}{\partial z_k(p)} \qquad \ldots 11$$

then Eq.10 yields:

$$\frac{\partial \varepsilon}{\partial w_{kj}} = -\Phi_k(p) x_j(p) = -\Phi_k y_j(p-1) \qquad \ldots 12$$

and, by Eqs.5 and 12

$$\Delta w_{kj} = \eta \Phi_k(p) x_j(p) = \eta \Phi_k(p) y_j(p-1) \qquad \ldots 13$$

j denoting the jth input to neuron k of the output (p) layer. Furthermore, by Eq. 11:

$$\Phi_k = -\frac{\partial \varepsilon}{\partial z_k} = -\frac{\partial \varepsilon}{\partial y_k} \frac{\partial y_k}{\partial z_k} \qquad \ldots 14$$

But, by Eq.2:

$$\frac{\partial \varepsilon}{\partial y_k} = -(d_k - y_k) = y_k - d_k \qquad \ldots 15$$

whereas, for a sigmoid nonlinearity:

$$y_k = F_N(z_k) = \frac{1}{1 + \exp(-z_k)} \qquad \ldots 16$$

therefore

$$\frac{\partial y_k}{\partial z_k} = y_k(1-y_k) \qquad \ldots 17$$

Consequently; by Eqs. 14, 15 and 17

$$\Phi_k = y_k(1-y_k)(d_k-y_k) \qquad \ldots 18$$

such that, at the output layer, by Eqs. 5 and 8:

$$\Delta w_{kj} = -\eta\frac{\partial \varepsilon}{\partial w_{kj}} = -\eta\frac{\partial \varepsilon}{\partial z_k}\frac{\partial z_k}{\partial w_{kj}} \qquad \ldots 19$$

Where by Eqs 9 and 14

$$\Delta w_{kj}(p) = \eta\Phi_k(p)y_j(p-1) \qquad \ldots 20$$

Φ_k being as in Eq.18, to complete the derivation of the setting of output layer weights.
Back-propagating to the rth hidden layer, we still have, as before

$$\Delta w_{ji} = -\eta\frac{\partial \varepsilon}{\partial w_{ji}} \qquad \ldots 21$$

for the ith branch into the jth neuron of the rth hidden layer. Consequently, in parallelity to Eq.8:

$$\Delta w_{ji} = -\eta\frac{\partial \varepsilon}{\partial z_j}\frac{\partial z_j}{\partial w_{ji}} \qquad \ldots 22$$

The $learning\ rate\ \eta$ should be adjusted stepwise, and noting Eq.9 and the definition of Φ in Eq.14:

$$\Delta w_{ji} = -\eta\frac{\partial \varepsilon}{\partial z_j}y_i(r-1) = \eta\Phi_j(r)y_i(r-1) \qquad \ldots 23$$

such that, by the right hand-side relation of Eq.14

$$\Delta w_{ji} = -\eta\left[\frac{\partial \varepsilon}{\partial y_j(r)}\frac{\partial y_j}{\partial z_j}\right]y_i(r-1) \qquad \ldots 24$$

Where $\frac{\partial \varepsilon}{\partial y_j}$ is inaccessible (as is, therefore, also $\Phi_j(r)$ above). However, ε can only be affected by upstream neurons when one propagates back- wards from the output. No other information is available at that stage. Therefore:

$$\frac{\partial \varepsilon}{\partial y_j(r)} = \sum_k \frac{\partial \varepsilon}{\partial z_k(r+1)}\left[\frac{\partial z_k(r+1)}{\partial y_j(r)}\right] = \sum_k \frac{\partial \varepsilon}{\partial z_k}\left[\frac{\partial}{\partial y_j(r)}\sum_m w_{km}(r+1)y_m(r)\right] \qquad \ldots 25$$

where the summation over k is performed over the neurons of the next $(the\ r+1)$ layer that connect to $y_j\ (r)$, whereas summation over m is over all inputs to each $k'th$ neuron of the $(r+1)$ layer. Hence, and noting the definition of Φ, Eq.25 yields:

$$\frac{\partial \varepsilon}{\partial y_j(r)} = \sum_k \frac{\partial \varepsilon}{\partial z_k(r+1)}w_{kj} = -\sum_k \Phi_k(r+1)w_{kj}(r+1) \qquad \ldots 26$$

Since only $w_{kj}\ (r+1)$ is connected to $y_j(r)$. Consequently, by Eqs.14, 17 and 26:

$$\Phi_j(r) = \frac{\partial y_j}{\partial z_j}\sum_k \Phi_k(r+1)w_{kj}(r+1) \qquad \ldots 27$$

$$= y_j(r)[1-y_j(r)]\sum_k \Phi_k(r+1)w_{kj}(r+1) \qquad \ldots 28$$

and, via Eq.20:

$$w_{ji}(r) = \eta\Phi_j(r)y_i(r-1) \qquad \dots 29$$

to obtain $\Delta w_{ij}(r)$ as a function of Φ and the weights of the $(r+1)$ layer, noting Eq.27.

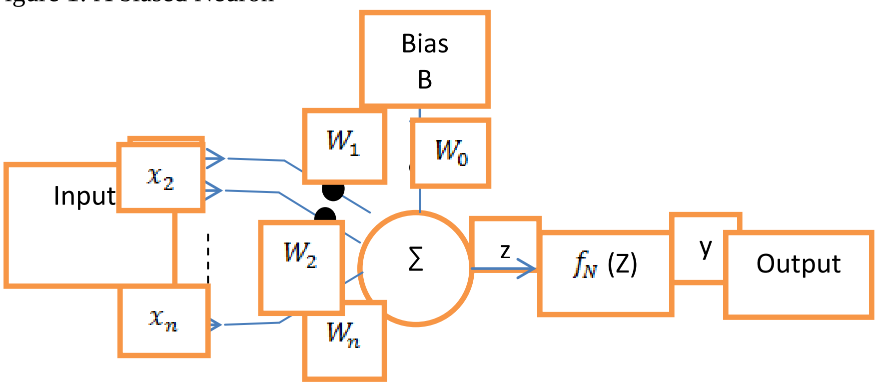
Introduction of bias into NN

It is often advantageous to apply some bias to the neurons of a neural network as presented in Figure 1. The bias can be trainable when associated with a trainable weight to be modified as is any other weight. Hence the bias is realized in terms of an input with some constant (say $+1$ or $+B$) input, and the exact bias b_i (at the ith neuron) is then given

$$b_i = w_{oi}B \qquad \dots 30$$

w_{oi} being the weight of the bias term at the input to neuron i .

Figure 1: A biased Neuron



Maximum Covariance Method

The proposed MC initialization method Mikko, Petri and Kimmon [1996] can be used to initialize MLPs with one hidden layer. The MC method can be directly expanded to multi-output case. The network considered can be written as

$$y = v_0 + \sum_{j=1}^{q} v_j \tanh\left(w_{0j} + \sum_{i=1}^{r} w_{ij}x_i\right) \qquad \dots 31$$

The number of inputs is r, number of hidden units is q, weights are denoted with v_j and w_{ij} (including the biases v_0 and w_{0j}), and the activation function in the hidden units is hyperbolic tangent $(tanh)$ function. It is noted that the output unit is linear. The RPROP training method, which is used after the initialization, can be expressed with the following equations

$$\theta(t+1) = \theta(t) + \Delta\theta(t) \qquad \dots 32$$

$$\Delta\theta(t) = \begin{cases} -\Delta(t) & , if\ \partial E^t/\partial\theta > 0 \\ +\Delta(t) & , if\ \partial E^t/\partial\theta < 0 \\ 0 & , else \end{cases} \qquad \dots 33$$

$$\Delta(t) = \begin{cases} \eta^+\Delta(t+1) & , if\ (\partial E^{t-1}/\partial\theta)(\partial E^t/\partial\theta) > 0 \\ \eta^-\Delta(t-1) & , if\ (\partial E^{t-1}/\partial\theta)(\partial E^t/\partial\theta) < 0 \\ \Delta(t-1) & , else \end{cases} \qquad \dots 34$$

Parameter θ denotes a weight (v_j or w_{ij}) and E is the cost function i.e. the sum squared error. The RPROP method includes several parameters for which we used the following values: decrease factor $\eta^- = 0.5$, increase factor $\eta^+ = 1.2$, initial update value $\Delta_0 = 10^{-5}$, maximum update value $\Delta_{max} = 1$ and minimum update value $\Delta_{min} = 10^{-10}$.

The maximum covariance initialization algorithm can be described by the following steps:

a. Choose the desired number of hidden units q by using some appropriate model selection method. Different model selection methods have been represented for example in [Lehtokangas 1995].

b. Create M candidate hidden units $(M \gg q)$ by initializing their weights w_{ij} with random values. We used $M = 10q$ and the candidate units were initialized with uniformly distributed random numbers from the interval $[-4; 4]$.

c. Do not connect the candidate units to the output unit yet. Only parameter feeding the output unit at this time is the bias weight v_0. Set the bias weight value to be such that the network output is the mean of the desired output sequence.

d. Calculate the covariance for each of the candidate unit from equation

$$C_j = \frac{1}{n}\sum_{p=1}^{n}(o_{j,p} - \bar{o}_j)(e_p - \bar{e}) \qquad , j = 1, \dots, M \qquad \dots 35$$

In which $o_{j,p}$ is the output of the jth hidden unit for pth pattern. Parameter $\bar{o}_j$ is the mean of the jth hidden unit's output, e_p is the output error at the network output and ê is the mean of the out errors.

e. Find the maximum absolute covariance $|C_j|$ and connect the corresponding hidden unit to the output unit. Set $M = M - 1$.

f. Optimize the currently existing output weights v_j with linear regression. Note that the number of these weights is increased by one every time a new candidate unit is connected to the output unit, and due to the optimization the output error changes each time.

g. If q candidate units have been connected to the output unit then quit the initialization phase; otherwise repeat the steps 3-5 for the remaining candidate units.

The idea behind the MC initialization method is to one by one select those hidden units amongst the candidates which have the maximum absolute covariance with the current output error. In this way those candidate hidden units are selected which can efficiently 'cancel' the output error.

4.0 Dataset and Experiments

Table 1 shows a small part of the data for our main example in Bradley, Trevor, Iain and Robert [2004]. Ten baseline variables, age, sex, body mass index, average blood pressure and six blood serum measurements, were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline. The statisticians were asked to construct a model that predicted response y from covariates $x_1, x_2, \dots, x_{10}$. Two hopes were evident here, that the model would produce accurate baseline predictions of response for future patients and that the form of the model would suggest which covariates were important factors in disease progression.

Let $x_1, x_2, \dots, x_m$ be m-vectors representing the covariates, m= 10 and n = 442 in the diabetes study, and let y be the vector of responses for the n cases. By location and scale transformations it is assumed that the covariates have been standardized to have mean 0 and unit length, and that the response has mean 0. The response Y is then class into 3 groups

Fasting Plasma Glucose (FPG):

Group1 = {normal up to 100 mg/dL} indicate no diabetes
Group2 = {between 100 and 125 mg/dL} indicate impaired fasting glucose or pre-diabetes.
Group3 = {126 mg/dL or higher} indicate diabetes.

or

Postprandial blood glucose test (PPB):
Group1 = {under 140 mg/dL.} indicate no diabetes
Group2 = {between 140 and 199mg/dL} indicate pre-diabetes.
Group3 = {200 and above value may} indicate diabetes.

Table 1: Diabetes study: 442 diabetes patients were measured on 10 baseline variables; a prediction model was desired for the response variable, a measure of disease progression one year after baseline

AGE	SEX	BMI	BP	S1	S2	S3	S4	S5	S6	Y
59	2	32.1	101	157	93.2	38	4	4.8598	87	151
48	1	21.6	87	183	103.2	70	3	3.8918	69	75
72	2	30.5	93	156	93.6	41	4	4.6728	85	141
24	1	25.3	84	198	131.4	40	5	4.8903	89	206
50	1	23	101	192	125.4	52	4	4.2905	80	135
23	1	22.6	89	139	64.8	61	2	4.1897	68	97
36	2	22	90	160	99.6	50	3	3.9512	82	138
66	2	26.2	114	255	185	56	4.55	4.2485	92	63
60	2	32.1	83	179	119.4	42	4	4.4773	94	110
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
36	1	30	95	201	125.2	42	4.79	5.1299	85	220
36	1	19.6	71	250	133.2	97	3	4.5951	92	57

Results

A confusion matrix [Kohavi and Provost 1998] contains information about actual and predicted classifications done by a classification system as presented in Figure 2. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a three class classifier.

Several standard terms have been defined for the 2 class matrix:

The accuracy (AC) is the proportion of the total number of predictions that were correct.

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation:

The false positive rate (FP) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

The true negative rate (TN) is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

The false negative rate (FN) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

Finally, precision (P) is the proportion of the predicted positive cases that were correct. From Figure 2, the diagonal cells show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red). The results show very good recognition.

Figure 2: Confusion Table of the Diabetes Mellitus Diagnosis Using the Network

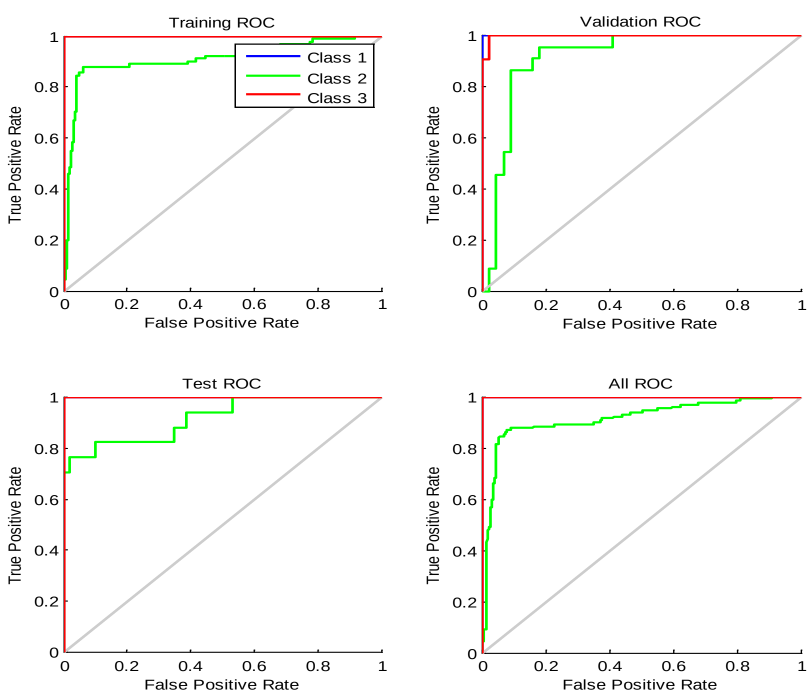


Figure 3: Receiver Operating Characteristic (ROC) curve

The colored lines in each axis represent the ROC curves. The ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) as the threshold is varied. A perfect test would show points in the upper-left corner, with 100% sensitivity and 100% specificity. For this problem, the network performs very well.

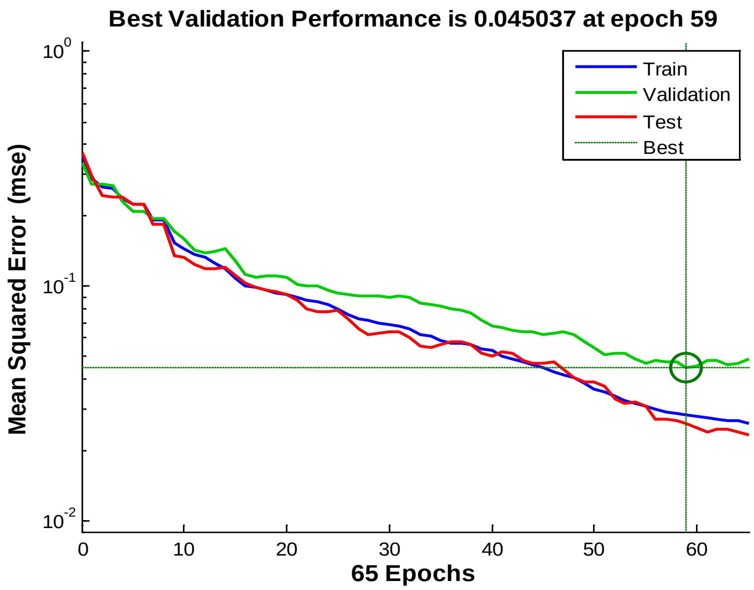


Figure 4: Validation Performance Curve

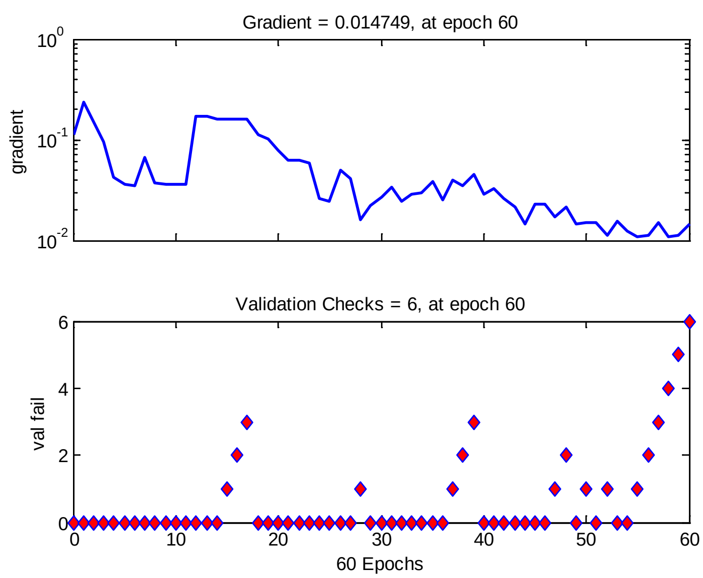


Figure 5: Gradient and Epochs

Conclusion

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified [Theodoridis and Koutroumbas 2006]. In this work, attempt is made at the development of a Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure. The Maximum covariance method is divided into three phases. A large number of candidate's hidden units was created by initializing their weights with random values. Then the desired number of hidden units is selected amongst the candidates by using the maximum covariance. The weights feeding the output units are calculated with linear regression. After the maximum covariance initialization, the network is trained with the resilient back propagation which is an adaptive training algorithm. The activation function in the hidden units is hyperbolic tangent function. The network is then trained 70%, tested (15%)and validated (15%) with ten (10) baseline variables, age, sex, body mass index, average blood pressure and six blood serum measurements, were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline. The result shows the algorithm is efficient in the diagnosis of how is a diabetic patient.

References

ALBERTI K.G, AND ZIMMET P.Z [1998]. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications Part 1: Diagnosis and Classification of Diabetes Mellitus. Diabet Med. 15(7): 539-53.

BARNETT, A.H., EFF, C., LESLIE, R.D.G. AND PYKE, D.A. [1981]. Diabetes in identical twins. Diabetologia, 20237-93.

BRADLEY EFRON, TREVOR HASTIE, IAIN JOHNSTONE AND ROBERT TIBSHIRANI [2004]. Least Angle Regression, *Annals of Statistics* (with discussion), 407-499.

DAVAR GIVEKI, HAMID SALIMI, GHOLAMREZA BAHMANYAR AND YOUNES KHADEMIAN [2012]. Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search. CORRABS/1202.3887. hhttp://arxiv.org/1201.2173. Accessed on 23rg August 2014.

GRAUPE DANIEL [2007]. Principles of Artificial Neural Networks (2nd Edition) Advanced Series on Circuits and Systems, Vol. 6. World Scientific Publishing Co. Pte. Ltd. ISBN-13 978-981-270-624-9.

HAYKIN S. [1999]. Neural Networks - A Comprehensive Foundation, 2nd ed., Prentice Hall.

IRVINE, WJ, TOFT, AD, HOLTON, DE, PRESCOTT, RJ, CLARKE, BF, AND DUNCAN, U P [1977]. Familial studies of type-I and type-II idiopathic diabetes mellitus. Lancet: 325-328.

KOHAVI, R., AND PROVOST, F. [1998]. On Applied Research in Machine Learning. In Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Columbia University, New York , Vol.30.

LEHTOKANGAS M. [1995]. Modeling with Layer Feedforward Neural Networks," Doctoral Thesis, Tampere University of Technology, Electronics Laboratory, Finland.

MCCULLUCH, W. S. AND PITTS, W. [1943] A Logical Calculus of the Ideas Imminent in Nervous Activity. Bulletin Mathematical Biophysics, Vol. 5, 116-132.

MIKKO LEHTOKANGAS, PETRI KORPISAARI AND KIMMO KASKI [1996]. Maximum Covariance Method for Weight Initialization of Multilayer Perceptron Networks. ESANN'1996 proceedings - European Symposium on Artificial Neural Networks. Bruges (Belgium), 24-25-26 April 1996, D-Facto public, ISBN 2-9600049-6-5, 243-248.

NATIONAL DIABETES DATA GROUP [1979]. Classification of Diabetes Mellitus and other Categories of

Glucose Intolerance. Diabetes Vol. 28:1039-1057.

POLAT, K., AND GUNES, S. [2007]. An Expert System Approach Based on Principal Component Analysis and Adaptive Neuro-Fuzzy Inference System to Diagnosis of Diabetes Disease. Digital Signal Processing, 17(4), 702–710.

ROSENBLATT, F. [1958]. The Perceptron, a Probabilistic Model for Information Storage and Organization in the Brain, Psychol. Rev. 65, 385-407.

ROSSANA M. S. CRUZ, HELTON M. PEIXOTO AND RAFAEL M. MAGALHAES [2011]. Artificial Neural Networks and Efficient Optimization Techniques for Applications in Engineering. In Artificial Neural Networks- Methodological Advances and Biomedical Applications, Edited by Kenji Suzuki, Published by InTech Janeza Trdine Vol. 9, 51000 Rijeka, Croatia

SILVA, P.H., DA F., CRUZ, R.M. S and D'ASSUNÇAO, A.G. [2010]. Neuromodeling and Natural Optimization of Nonlinear Devices and Circuits, In: System and Circuit Design for Biologically-Inspired Intelligent Learning, Turgay Temel (Ed.), IGI Global, ISBN 9781609600181, Hershey PA. No prelo.

SILVA, P.H., DA F., CRUZ, R.M. S and D'ASSUNÇAO, A.G. [2010a]. Blending PSO and ANN for Optimal Design of FSS Filters with Koch Island Patch Elements, IEEE Transactions on Magnetics, Vol. 46, No. 8.

THEODORIDIS SERGIOS AND KOUTROUMBAS KONSTANTINOS [2006]. Pattern Recognition Elsevier USA

WILLIAM C. KNOWLER, DAVID J. PETTITT, PETER H. BENNETT, AND ROBERT C. WILLIAMS [1983]. Diabetes Mellitus in the Pima Indians: Genetic and Evolutionary Considerations. American Journal of Physical Anthropology 62:107-114.