

*Research Article***A Comprehensive Evolution for Applicability of Machine Learning Algorithms on Various Domains****Zuleyha Akusta Dagdeviren** *International Computer Institute, Ege University, Izmir, Turkey*

ARTICLE INFO

Article history:

Received 1 October 2020

Accepted 21 November 2020

*Keywords:*Color reduction
Gradient descent
Linear regression
Logistic regression
Machine learning
Self-organizing map

ABSTRACT

Machine learning algorithms are able to learn from data, make decision and improve what they learn by having experience without human intervention. Machine learning techniques are becoming increasingly important nowadays that everything is going to be fully automated. They have been used in various fields such as recommendation engines, self-driving cars, offering personal suggestions from retailers, cyber fraud detection, face recognition, and etc. This study presents some of the most commonly used machine learning techniques from supervised and unsupervised learning classes such as linear regression, logistic regression, neural networks, and self-organizing map. In linear regression technique, it is tackled to fit a linear function to user data in order to model the relationship between variables. It can be a useful technique to make weather estimation, to understand marketing effectiveness and to model consumer behavior. Logistic regression is a statistical model that uses a logistic function and is appropriate when dependent variable is binary. Neural networks mimic the operation of human brain to recognize patterns from the underlying data. They have wide range of application such as cancer diagnosis, e-mail spam filtering, and signal classification. Self-organizing map, a special type of neural networks, is utilized to achieve dimensionality reduction that generally used for seismic analysis, project prioritization, and image processing such as color reduction. Each implementation in this study shows that the success of the results obtained by applying machine learning techniques depends on using the right technique in the appropriate area.

This is an open access article under the CC BY-SA 4.0 license.
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

Machine learning is a type of artificial intelligence that can achieve results that are not explicitly programmed and focuses on the computer software which learns to grow and develop on their own when disclosed new data. In recent years, it has become one of the main pillars of information technologies and has become a part of our live. With the increasing amount of data, intelligent data analysis is becoming more common as an essential component of technological progress [1,2].

With wide-ranging applications, machine learning is one of the significant and spreading areas of computer science. Machine learning techniques are successfully used to solve various problems in image processing, pattern recognition, robotics, and etc. [3]. Machine learning is the basis of numerous important applications

such as self-driving cars, email anti-spam, speech recognition, face recognition, web search, social networking, product recommendations and advertising [4].

Supervised and unsupervised learning are two main classes of machine learning methods. In supervised learning that requires an exterior teacher, each output is taught what the appropriate answer should be. In other words, supervised learning aims to gain an understanding of input to output mapping with suitable values fed by a supervisor. During the learning process, global information may be required. Supervised learning uses classification algorithms and regression methods. Neural networks, linear regression, logistic regression can be given as examples of supervised learning.

Unsupervised learning does not use such external supervisor and is based upon only local information. It

uses unknown and unlabeled training data and also the number of clusters is unknown. The aim is to discover certain patterns in input data to see how it usually behaves. Self-organizing map and k-means are kinds of unsupervised learning methods.

In this study, several supervised and unsupervised machine learning algorithms are investigated both in theoretical and practical aspects. For each algorithm, we first present its theoretical foundation in brief then we apply it on an example domain to show its effectiveness. We also comment on the time consumption of the algorithms. Firstly, we study on linear regression methods including brute force, gradient descent, normal equation and their implementations on covid-19 coronavirus data. Following this, we investigate logistic regression and its application which is a feature classification example applied on segmented images. Moreover, we work on self-organizing map and k-means algorithms and their implementations for color reduction.

The rest of the paper is organized as follows: Linear regression methods with its covid-19 coronavirus data applications are presented in Section 2. Logistic regression and its usage in image segmentation are given in Section 3. Section 4 includes the self-organizing map and k-means algorithms which are implemented for color reduction. The paper ends with the conclusion in Section 5.

2. Linear Regression Methods

Machine learning techniques analyze the association of two variables. For each instance, the hypothesis line fitting through the training set is aimed to determine and these variables are known [5]. When the relationship between a quantitative result and a single quantitative explanatory variable are examined, the most commonly considered analysis method is linear regression [6-9]. Linear regression is represented with the Equation 1.

$$h_{\theta}(x) = \theta X = \theta_0 x_0 + \theta_1 x_1 \tag{1}$$

In this formula, the coefficients of θ_0 and θ_1 are x_0 and x_1 , respectively. The coefficient x_0 is known as Bayes input and always equals to 1 [10]. So, the hypothesis function becomes the form of Equation 2.

$$h_{\theta}(x) = \theta_0 + \theta_1 x \tag{2}$$

The function of the parameters θ_0 and θ_1 called as cost function $J(\theta_0, \theta_1)$, is given in the Equation 3 where m is the number of training examples. The aim is to minimize the cost function.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2 \tag{3}$$

The row that gives the minimum summation is determined, so that the values corresponding to the same

row in θ matrix are obtained. By the values of θ_0 and θ_1 in this row, the hypothesis function $h(x)$ is constructed.

Table 1. 20 Days New Covid-19 Cases and New Deaths in US [11]

Day (April,2020)	New Cases	New Deaths
1	26,942	1,243
2	30,413	1,182
3	32,769	1,263
4	34,517	1,545
5	25,753	1,409
6	31,601	1,505
7	33,877	2,228
8	32,344	2,165
9	33,946	2,111
10	34,089	2,236
11	30,338	2,024
12	27,766	1,727
13	26,985	1,726
14	27,260	2,566
15	30,521	2,631
16	29,980	2,193
17	32,368	2,543
18	29,079	1,883
19	26,113	1,570
20	28,131	1,952

In this section, various popular linear regression methods which are brute force, gradient descent and normal equation are applied on Covid-19 coronavirus dataset given in Table 1. This data taken from [11], includes the number of people caught with coronavirus and died from coronavirus in United States within the first 20 days of April 2020.

2.1. Brute Force Method

One of the linear regression methods that warrants to produce an exact solution is brute force hence it tries set of all possible candidates and may produce optimal solutions in case of small datasets are given. On the other side, its execution time generally reaches an exponential order of growth for larger datasets. In most practical settings, brute force may be unacceptable for tackling with problems, since the execution time may span many years [12].

When brute force method is applied for the prediction of new infected patients and new deaths from Covid-19, hypothesis function is iteratively calculated with different θ values θ_0, θ_1 . To change these θ value couples incrementally, range is intuitively defined as [-3000 3000]. θ matrix is constructed as seen below:

$$\theta = [\theta_0 \ \theta_1] = \begin{bmatrix} -3000 & -3000 \\ -3000 & -2990 \\ \vdots & \vdots \\ 3000 & 2990 \\ 3000 & 3000 \end{bmatrix}$$

Then, for each i^{th} row contains θ parameters pair, the result matrix that contains estimated y values is constructed by the Equation 2. The differences between y value and

estimated y value is calculated and the sum of the differences in the same row is calculated. Hypothesis function is created with the coefficients θ_0 and θ_1 in theta matrix line corresponding to the line in result matrix giving the closest results to real new cases value. The number of people caught Covid-19 are calculated as 24500 with the below theta values.

$$\theta_{cases} = \begin{bmatrix} 34500 \\ -500 \end{bmatrix}$$

$$h_{\theta}(20) = 34500 + (-500) \times 20 = 24500$$

By the hypothesis function, the number of people died from Covid-19 of 20th April is predicted as 3170. The result is not good enough because learning rate is taken too large and the brute force algorithm overshoots the minimum. If it is defined as smaller, better result can be obtained but runtime of the algorithm becomes too long to be implemented.

$$\theta_{deaths} = \begin{bmatrix} 970 \\ 110 \end{bmatrix}$$

$$h_{\theta}(20) = 970 + 110 \times 20 = 3170$$

2.2. Gradient Descent

Gradient descent method that is an optimization technique for determining the local minimum. This method is very popular among scientists because it is straightforward and terminates by executing relatively less step count. For the hypothesis function $h(x)$, that is differentiable in a given limit, the direction of its fastest decline is the negative gradient. Gradient descent algorithm is applied to detect local minimum [13].

In the gradient descent algorithm, θ values are calculated by the Equation 4 for j equals to 0 and 1, separately. The algorithm determines the right θ values, by changing θ_0 , θ_1 values to reduce cost function given in Equation 3 until ending up at a minimum cost. For each step of gradient descent all the training examples are used.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \tag{4}$$

If learning rate α is too large, gradient descent algorithm goes past the minimum unintentionally, if it is too small then gradient descent works slowly. Firstly, θ values are initialized by 0 and α is defined as 0.01. Regarding to the Equation 4, θ values are determined and then the hypothesis function given in Equation 2 is constructed.

Gradient descent algorithm is implemented on coronavirus dataset to predict the number of people infected with Covid-19 on 20th April. The θ parameters are determined as seen below by the Equation 4. Regarding to these parameters the hypothesis function is constructed and the number of new cases is predicted as 30746.862.

$$\theta_{cases} = \begin{bmatrix} 28500.91 \\ 112.30 \end{bmatrix}$$

$$h_{\theta}(20) = 28500.91 + 112.30 \times 20 = 30746.862$$

The hypothesis line is shown on the diagram of new cases in Figure 1. The difference between the actual value and the calculated value is not small enough but acceptable because the dataset size is not large enough for the system to learn well.

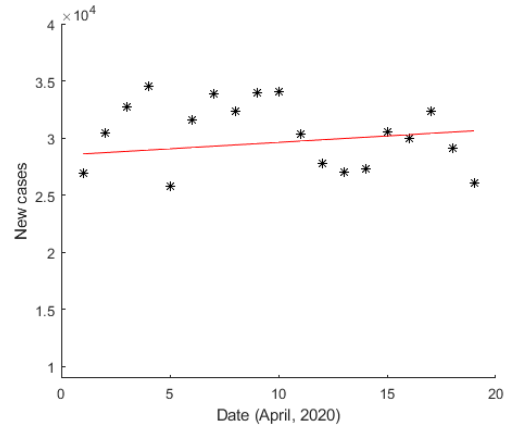


Figure 1. New covid-19 cases vs. date diagram with the hypothesis line

Reported deaths from covid-19 coronavirus are estimated by the gradient descent as 2469.139 with θ values given below. The number of the reported deaths on 20th April is 1952 as seen on the Table 1. The estimated value is not close to the real value due to the smallness of the dataset. The reported deaths and the hypothesis line are shown in Figure 2.

$$\theta_{deaths} = \begin{bmatrix} 1231.447 \\ 61.885 \end{bmatrix}$$

$$h_{\theta}(20) = 1231.447 + 61.885 \times 20 = 2469.139$$

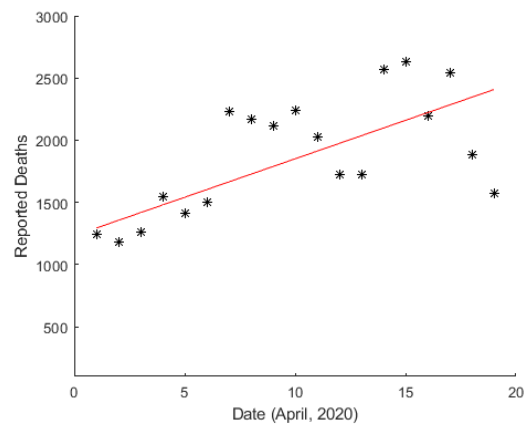


Figure 2. Reported deaths from covid-19 vs. date diagram with the hypothesis line

2.3. Normal Equation

The normal equation method is an approach to fitting a mathematical model to data in cases where the idealized value provided by the model for any data point is expressed linearly in terms of the unknown parameters of the models [14].

Normal equation is the set of equations arising in the least squares method whose solutions give the constants

that determine the shape of the estimated function [15]. This method solves the Θ matrix analytically by the Equation 5 in matrix notation [16]. The hypothesis with these Θ values minimizes the cost function in Equation 3 or squared error in other words [17].

$$\theta = (X^T X)^{-1} X^T y \tag{5}$$

Using 19-day data given in Table 1, number of new cases infected by covid-19 and new deaths from coronavirus on 20th April are estimated by normal equation. For 20th April, the estimated number of infected new cases is calculated by hypothesis function with obtained Θ values as below:

$$\theta_{cases} = \begin{bmatrix} 31669 \\ -131.8 \end{bmatrix}$$

$$h_{\theta}(20) = 31669 + (-131.8)20 = 29033$$

Θ matrix for new reported deaths due to covid-19 on 20th April is obtained as seen below. The number of the deaths is calculated as 2395.308 by the Equation 2 where the real value is 28131 as seen in Table 1.

$$\theta_{deaths} = \begin{bmatrix} 1367.8 \\ 51.375 \end{bmatrix}$$

$$h_{\theta}(20) = 1367.8 + 51.375 \times 20 = 2395.308$$

The results obtained by normal equation are better than gradient descent algorithm. When the iteration number is increased in the gradient descent algorithm, there is no significant difference and almost the same results are obtained with normal equation but runtime increases.

3. Logistic Regression

Logistic regression method is appropriate to make predictive analysis when the dependent variable is binary. In this method variables are categorized such as win/loss, alive/dead, healthy/sick or success/failure and so on [18]. Hence, it can be named as binary classification. For example, to predict whether a tumor is malignant or not, logistic regression can be used. It is also used to determine the relationship between one dependent binary variable and interval, ordinal or nominal independent variables. This method can be extended to model multiclass classification such as determining whether an image includes a person, flower, dog, and etc. Every object detected in the image is assigned a probability between 0 and 1 [19].

Image Segmentation labeled dataset [20] from the UCI Machine Learning Repository is used as training dataset for this part. Segmentation is to subdivide an image into its component objects or regions [21]. The Image Segmentation instances were drawn randomly from a database of real seven outdoor images. The images were handsegmented to create a classification for each pixel. Each pattern has 19 continuous attributes and corresponds to a 3x3 region of an outdoor image. There are 210 patterns

in the training set and 2100 patterns in the test set. There are 300 patterns from each class in test data.

Classification includes assignment of data points to known classes by using supervised learning. Logistic regression algorithm is applied on the Image Segmentation dataset to classify the patterns into one of the seven classes which are predefined as brickface, sky, foliage, cement, window, path and grass.

During dataset processing, the third attribute, that is region-pixel-count: the number of pixels in a region, is not considered because it has the same value as $3 \times 3 = 9$ for each data set. There are 18 features after removing third feature. The data is normalized to the range of [-5 5].

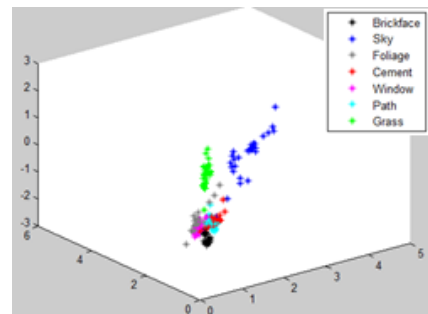


Figure 3. The graph of the classes with three features

The logistic regression hypothesis cannot be defined regarding to the large features. To use less features, image segmentation data is observed by plotting the training data with three features. The classification result obtained by logistic regression with three features (attributes 11, 12 and 16) is shown in Figure 3 that each data point is assigned to one of the seven classes.

As seen from Figure 3, some data points from different classes are overlapping. To avoid this issue, classification can be done with two features. So, the logistic regression is implemented on dataset with two features that are experimented by every pair combination out of three attributes. As a result of experiments, logistic regression is applied with attributes 11 and 12. The classified data points are shown in Figure 4.

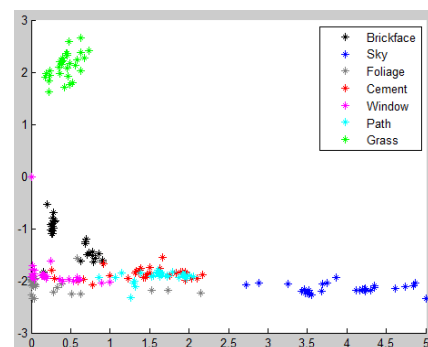


Figure 4. The graph of the classes with two features

4. Self-Organizing Map for Color Reduction

All algorithms mentioned so far fall into the supervised learning in which labeled training data assumed to be available. In this section, some important machine learning algorithms that are trained by using unsupervised learning are introduced and a sample color reduction implementation in which they are used together is given.

Neural networks are used for applications where formal analysis would be difficult or impossible, such as pattern recognition and nonlinear system identification and control. The following section focuses on a particular type of neural network model which is known as self-organizing map [22].

4.1. Self-Organizing Map

The artificial neural network (ANN) was first proposed by Teuvo Kohonen in 1982 [23]. ANNs inspired by biological nervous systems are aimed to model the information processing paradigm [22]. A self-organizing map (SOM) that is trained using unsupervised learning approach is a special type of ANN.

A given input data pattern; SOM has the capability to classify data itself without any external supervision. The training process in SOM is competitive. The basic components are neurons which allow it to take meaningful decisions [24]. It is also a dimension reduction technique that uses a mapping or projection procedure from a large dimension to a smaller dimension. It helps in reducing the dimensions of data by grouping the input patterns that have a similarity among themselves [25].

4.2. K-Means

One of the most widely applicable clustering methods is k-means [26]. This algorithm is used frequently in many areas such as customer segmentation, market analysis, computer vision and etc. K-means algorithm applied by selecting a convenient starting strategy acts as a very efficient color reduction method [27].

After the number of clusters k value is determined, k center points are selected randomly without replacement. For each input data point, the distance between randomly determined center points is calculated. The data is assigned to a cluster based on its closest center point. Then center points are determined again for each cluster and clustering is performed according to the new center points. These steps continue until the system becomes stable.

4.3. Color Reduction

The image color reduction is an essential technique used in many applications, such as compression, presentation, transmission, segmentation and analysis of color images [27-29]. For a defined particular data type, the RGB color cube which is seen in Figure 5, is a 3 dimensional array that includes all the colors of the image. Three color cubes are defined for RGB images in MATLAB. For an RGB image in uint8 type class, there are

256 pixel values for each color plane. Hence, there are 224 colors in the color cube. Regardless of what colors are used in the image, the same color cube applies to all uint8 RGB images.

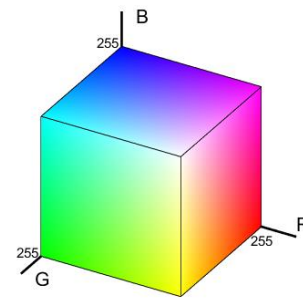


Figure 5. RGB color cube for unit8 (256 colors) image

Color reduction is a method that firstly the color cube is partitioned into smaller cubes. Then each color is matched to the color value at the center of the small cube that it corresponds.

The goal of the color reduction is to minimize disparity between the colors of the original image and the quantized colors. Our aim is to decrease inconsistencies between the actual and reduced colors. Figure 6 shows the color reduction of a uint8 image that is obtained by taking a slice of RGB color cube.

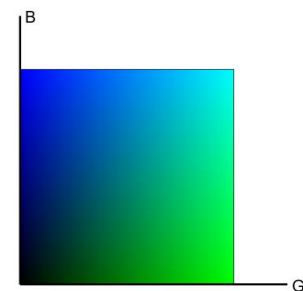


Figure 6. A slice of RGB color cube

SOM is one of the most effective color reduction methods and generally produces desired results [29,30-33]. In SOM algorithm, the initial weights can be assigned randomly or can be chosen from data randomly or these values can be defined by k-means method. In this study, SOM is applied with initialization by k-means algorithm. SOM based color reduction method is applied on 64 colors parrots image seen in Figure 7.a.

The algorithm generates the color reduced images obtained by different number of colors that are also the number of clusters. After color reduction, the output images constituted of 32 colors, 16 colors and 8 colors are seen in Figure 7.b, c and d, respectively. The experimental results show that, combining SOM with k-means algorithm provides an effective color reduction that the dominant colors of the input image retain even if the number of colors is reduced to 8.

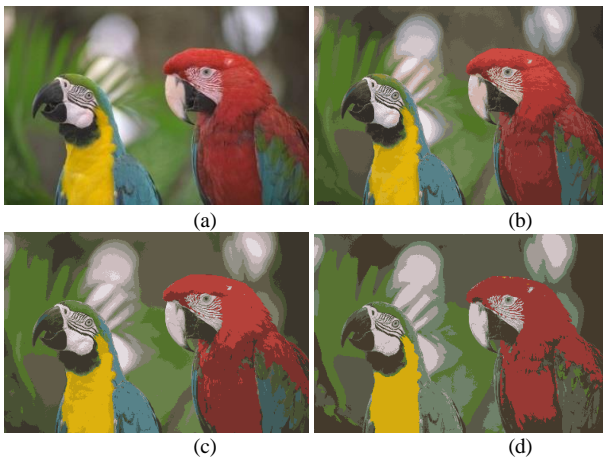


Figure 7. a) Original image b) 32 colors image c) 16 colors image d) 8 colors image

5. Discussion and Conclusions

In this article, we explain machine learning techniques using examples to make them as simple as possible. Brute force is a straightforward approach directly based on problem statement and definitions. This approach guarantees for finding the solution and prefers simpler hypothesis, but it is not effective for large data. Applying normal equation as a linear regression method for computing Θ provides some advantages such as no need to choose learning rate and not required to make many executions. This technique is not good for large data because when the matrix size is large, the algorithm works slowly. The gradient descent method requires selection of α and needs much iteration but works well even if data is

large. Logistic regression is one of the simple ways to implement binary classification and can be extended to multiclass classification. This method allows models to be updated easily to adopt new incoming data, unlike support vector machines or decision trees.

Neural network approach having ability of capturing complex relationships is suitable for large training set. Even if training data has many features, the network can learn the system. SOM technique, which is a type of neural network, provides easy data mapping. Because of its suitability for data compression, it is a preferable method for color reduction. The implementation of SOM-based color reduction in this paper, produces acceptable results.

We clarify that the effectiveness of the machine learning techniques depends on domain which they are applied. In another words, this study shows that selecting the appropriate machine learning method to tackle with a given problem domain is of utmost importance. Machine learning techniques studied in this paper, their advantages, disadvantages and possible domains that each method giving better results are summarized in Table 2.

As a future work, distributed machine learning techniques [34, 35] that provide parallelization of the training process and scale to larger datasets will be considered. We also plan to study on deep learning one of the ways of executing machine learning. In this manner, we aim to use graph convolutional networks to solve NP-Hard graph theoretical problems such as vertex cover, dominating set and independent set which are popular and have wide application areas.

Table 2. A Summary of Machine Learning Techniques

Method	Advantages	Drawbacks	Possible Domains
Brute Force	-No need to domain-specific knowledge -Straightforward approach -Finds exact solution	-Exponential time for execution -Not suitable for problems having a hierarchical structure	-Small-scale datasets -Smaller instances or modules of a larger dataset
Gradient Descent	-Very easy to use	-Needs more iterations hence has long execution time -Need to choose learning rate	-Datasets with very large number of features (N) -Small/Large datasets
Normal Equation	-No need to choose learning rate -No need to iteration	-Slower for even modest sized datasets -Numerically unstable	-If N is relatively small (on the order of a hundred ~ ten thousand)
Logistic Reg.	-Easy to implement -Can be extended to multiclass classification	-Nonlinear problems cannot be solved -Difficult to capture complex relationships	-Large-scale datasets -Dataset having features that are linearly separable
SOM	-Data mapping is easily understood and interpreted	-Difficult to determine input weights -Mapping can result in divided clusters	-Image dataset -Data compression -Anomaly detection
K-means	-Simple and flexible to implement -Scales to large dataset	-Dependent on initial values -Choosing k value manually	-Unlabeled data -Numerical data with a relatively lower number of dimensions

Acknowledgment

The author would like to thank Assoc. Prof. Dr. Muhammed Cinsdikici for his guidance about the algorithms used in this paper.

Author's Note

Abstract version of this paper was presented at 9th International Conference on Advanced Technologies (ICAT'20), 10-12 August 2020, Istanbul, Turkey with the title of "A Comprehensive Evolution for Applicability of Machine Learning Algorithms on Various Domains".

References

- [1] Shalev-Shwartz, S., & Ben-David, S., *Understanding Machine Learning: From Theory to Algorithms*, Cambridge: Cambridge University Press, 2014.
- [2] Smola, A. and Vishwanathan, S.V.N., *Introduction to Machine Learning*, Cambridge University Press, 2008.
- [3] Alpaydin, Ethem, *Introduction to Machine Learning*, Third Edition, MIT Press, 2014.
- [4] Ng, Andrew, *Machine Learning Yearning*, 2018.
- [5] Mohan, V. and Gupta, M., Data Regression with Normal Equation on GPU using CUDA, *International Journal of Computer Science, Information Technology and Security (IJCSITS)*, Vol. 2, No.2, April 2012.
- [6] Monahan, J. F., *A Primer on Linear Models*, CRC Press, 2008.
- [7] Rao, C.R., Toutenburg, H., Shalabh, and Heumann, C., *Linear Models and Generalizations - Least Squares and Alternatives*, Springer, 2008.
- [8] Khuri, Andre I., *Linear Model Methodology*, CRC Press, 2010.
- [9] Seltman, H. J., *Experimental Design and Analysis*, 2018 Available from: <http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf>
- [10] Montgomery, Douglas C., Peck, Elizabeth A., Vining, G. Geoffrey, *Introduction to Linear Regression Analysis*, Fifth Edition, John Wiley & Sons, Inc., 2012.
- [11] <https://www.worldometers.info/coronavirus/country/us/>
- [12] Wilson, R.A. and Keil, F.C., *The MIT Encyclopedia of the Cognitive Sciences*, 1999.
- [13] Toussaint, M., Lecture Notes: Some notes on gradient descent, Berlin, May 3, 2012.
- [14] Wang, X., Method of Steepest Descent and its Applications, 2008.
- [15] The Free Dictionary [Internet]. "Normal Equations". McGraw-Hill Dictionary of Scientific & Technical Terms, 6E, The McGraw-Hill Companies, Inc., 2003 [cited 27 Apr. 2020]. Available from: <https://encyclopedia2.thefreedictionary.com/Normal+Equations>
- [16] Rawlings, J. O., Sasthy, G. P., Dickey, D. A., *Applied Regression Analysis: A Research Tool*, Second Edition, 8 Springer-Verlag New York, Inc., 1998.
- [17] Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, Third Edition, Prentice Hall, 2010.
- [18] Kaplan, D. T., *Statistical Modelling A Fresh Approach*, E-book version of Second Edition, Project MOSAIC Books, 2017.
- [19] Sperandei, S., Understanding logistic regression analysis, *Biochemia medica*, 24. 12-8. 10.11613/BM.2014.003, 2014.
- [20] Dua, D. and Graff, C., UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science, 2019. Available from: <http://archive.ics.uci.edu/ml>
- [21] Gonzalez, R.C. and Woods, R.E., *Digital image processing*, Third Edition, Upper Saddle River, N.J.: Prentice Hall, 2008.
- [22] Rojas, R., *Neural Networks: A Systematic Introduction*, Springer-Verlag, Berlin, 1996.
- [23] Kohonen, T., Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, vol.66, pp.59–69, 1982.
- [24] Kohonen, T., *Self-Organizing Maps*, Springer-Verlag, Berlin, 2001.
- [25] Chetia, S. and Sarma, K. K., PCA and SOM based Dimension Reduction Techniques for Quaternary Protein Structure Prediction, *International Journal of Computer Applications* (0975 – 8887), Volume 46– No.18, May 2012.
- [26] Gan, G., Ma, C. and Wu, J., *Data Clustering: Theory, Algorithms, and Applications*, SIAM, 2007.
- [27] Celebi, M.E., Improving the performance of k-means for color quantization, *Image and Vision Computing*, vol.29, pp. 260–271, 2011.
- [28] Atsalakis, A. and Papamarkos N., Color reduction by using a new self-growing and self-organized neural network, *Vision, Video and Graphics*, 2005.
- [29] Zagoris, K., Papamarkos1, N. and Koustoudis, I., Color Reduction Using the Combination of the Kohonen Self-Organized Feature Map and the Gustafson-Kessel Fuzzy Algorithm, P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 703–715, Springer-Verlag Berlin Heidelberg, 2007.
- [30] Chang, C.H., Xu, P., Xiao, R. and Srikanthan, T., New adaptive color quantization method based on self-organizing maps, *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 237–249, 2005.
- [31] Hu, Y.-C., Lee, M.-G., K-means based color palette design scheme with the use of stable flags, *Journal of Electronic Imaging*, vol. 16 (3), 2007.
- [32] Hu, Y.-C. and Su, B.-H., Accelerated k-means clustering algorithm for colour image quantization, *Imaging Science Journal*, 56 (1) pp. 29–40, 2008.
- [33] Park, H.J., Kim, K.B., and Cha, E.-Y., An Effective Color Quantization Method Using Color Importance-Based Self-Organizing Maps, *Neural Network World*, 121-137, 2015.
- [34] Peteiro-Barral, D. and Guijarro-Berdiñas, B., A survey of methods for distributed machine learning. *Progress in Artificial Intelligence* 2, 1 (2013), 1–11, 2013.
- [35] Qiu, J., Wu, Q., Ding, G., Xu, Y. and Feng, S., A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, 1 (2016), 67, 2016.