Prediction of Breast Cancer by the Light weight Systems

K. Mahesh Kumar*, KVS Sivakrishna

MS (Software Engineering), Vit Univeristy, Vellore.

*Corresponding author: E-Mail: kokanti.maheshkumar2013@vit.ac.in

ABSTRACT

In the era of smart technology, from weather news to online order is known with the help of applications in the light weight systems. Through this paper we developed a method which helps in prediction of breast cancer of a patient by the product metric evaluation. By using this method, we reduce the characteristics required to predict the breast cancer in the light weight systems which has less computation power.

KEY WORDS: Data Analysis, Pearson, Regression.

1. INTRODUCTION

To analyse the dataset of breast cancer of various patients firstly, we have applied Pearson (A) method (Hauke, 2011). Pearson helps us to discover how intently a trait is related with different qualities. We have a dataset of almost 300 patients portraying their cell condition with the assistance of 9 characteristics. Among these we have selected sample of 13 patients. In light of the examination of data we can choose which quality can be considered and which property can be disregarded. For example, in Pearson strategy if the estimation of r is more than 0.8 then the traits are thought to be emphatically related and on the off chance that it is beneath 0.8 the traits are scantily related. Secondly we have applied Linear regression model, is used for demonstrating the relationship between a scalar dependent variable y and at least one systematic factors meant by x (Rashad Manzoor, 2017).

Table.1. Values

Thickness A1	Tissue size A2	Shape.uni A3	Adhesion A4	Epi A5	Bare A6	Bland A7	Normal A8	Mitosis A9
5	1	1	1	2	1	3	1	1
5	4	4	5	7	10	3	2	1
3	1	1	1	2	2	3	1	1
6	8	8	1	3	4	3	7	1
4	1	1	3	2	1	3	1	1
8	10	10	8	7	10	9	7	1
1	1	1	1	2	10	3	1	1
2	1	1	1	2	1	3	1	1
2	1	1	1	2	1	1	1	5
4	2	2	1	2	1	2	1	1
1	1	1	1	1	1	3	1	1
2	1	1	1	2	1	2	1	1

Pearson:

- Pearson based Attribute Clustering
- Create a table between the different characteristics as Appeared in IV A
- Complete the table utilizing augmentation of trait qualities
- Calculate the aggregate of every segment independently
- Substitute every one of the traits in the equation offered

Underneath to get the Pearson coefficient.

$$r = \frac{N \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{(N \cdot \sum x^2 - (\sum x)^2) (N \cdot \sum y^2 - (\sum y)^2)}}$$

Here N is the quantity of sets of scores;

$\sum xy$ Is the entirety of the result of matched scores

$\sum x$ is the entirety of x scores.

$\sum y$ is the entirety of y scores

$\sum x^2$ is the entirety of squared of x scores

$\sum y^2$ is the entirety of squared of y scores

Each of the relations with r under 0.8 is considered as week relations and the related characteristics can be overlooked and just emphatically related qualities are considered. This reduces the general number of qualities.

Table.2. Pearson Relationship

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	1	0.85	0.84	0.81	0.85	0.61	0.85	0.84	0.38
A2	0.85	1	0.99	0.79	0.81	0.69	0.8	0.96	0.24
A3	0.84	0.99	1	0.79	0.81	0.69	0.8	0.96	0.24
A4	0.81	0.79	0.79	1	0.91	0.76	0.87	0.71	0.27
A5	0.85	0.81	0.81	0.91	1	0.84	0.83	0.72	0.4
A6	0.61	0.69	0.69	0.76	0.84	1	0.72	0.62	0.23
A7	0.85	0.80	0.80	0.87	0.83	0.72	1	0.79	0.33
A8	0.84	0.96	0.96	0.71	0.72	0.62	0.79	1	0.27
A9	0.38	0.24	0.24	0.27	0.4	0.23	0.33	0.27	1

In light of the Pearson relationship property examination of breast cancer dataset which appears in the above Table II, taking the r value as 0.8, it is evident that the discoveries give how dependence of the attribute x on y decreases for various attributes

Linear Regression Method:

- Make a table between the different properties as shown in IV B
- Fill in the estimations of x and y
- Fill in the table with $h_{\theta}(x^i)$ values
- Fill the table with $h_{\theta}(x^i)-(y^i)$ values
- Fill the table with $(h_{\theta}(x^i)-(y^i))^2$ values
- Supplant the total values in the accompanying structure to get the enhanced cost of 0

$$J(\theta_1) = 1/2m \sum (h_{\theta}(x^i) - (y^i))^2$$

Table.3. Linear Regression

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	1	1.9, 1	2.0, 1	1.2, 0.5	1.4, 0.5	-	1.3, 1	1.3, 1	-
A2	1.9, 1	1	0.1, 1	-	1.8, 0.5	-	2, 1	0.6, 1	-
A3	2.0, 1	0.1, 1	1	-	1.8, 0.5	-	2, 1	0.6, 1	-
A4	1.2, 0.5	-	-	1	0.7, 1	-	1.2, 1	-	-
A5	1.4, 0.5	1.8, 0.5	1.8, 0.5	0.7, 1	1	2.9, 0.5	1.3, 1	-	-
A6	-	-	-	-	-	1	-	-	-
A7	1.3, 1	2.0, 1	2.0, 1	1.3, 1	-	-	1	-	-
A8	1.3, 0.5	0.6, 1	0.6, 1	-	-	-	-	1	-
A9	-	-	-	-	-	-	-	-	1

In light of Linear Regression (Fioravanti, 2000), we will get the enhanced θ value, and for this θ trait value can be proficiently relapsed after the Pearson likeness measure (where $r > 0.8$). This technique is utilized for cost improving which implies advancing the variable. This is one of the strategies utilized for accomplishing machine learning. Table.3, comprises of θ incentive alongside cost function for each quality connection. Attributes having same θ can be formed in to the same cluster.

2. CONCLUSION

By applying both Pearson and regression analysis we reduce the no of characteristics required to predict the breast cancer. If the patient submits his data, based on the formed clusters of previous patient's data, new patient's data will be evaluated. The light weight systems consist of less computing power rams, which is not possible to evaluate the data if more characteristics are there, so reduction in the characteristics would make the system to predict the breast cancer with less computing power.

REFERENCES

- Dhavachelvan P, Chandra Segar T, Satheskumar K, Evaluation of SOA Complexity Metrics Using Weyuker's Axioms, IEEE Int. Ad.Comp. (IACC), India, 2009, 2325 – 2329.
- Fioravanti F, Nesi P, A method and tool for assessing object-oriented projects and metrics management, Jnl. of Systems and Software, 53 (2), 2000, 111-136.
- Hauke J, Kossowski T, Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of Data, Quaestiones Geographicae Bogucki Wydawnictwo Naukowe, 30 (2), 2011, 87–93.
- Rashad Manzoor, Chandrasegar Thirumalai, Analysing the Breast Cancer Tissue using Semi-Supervised Learning And Similarity Measure, IEEE IPACT, 2017.
- Yule G.U, On the significance of Bravais' formulae for regression, in the case of skew correlation, Proceedings of the Royal Society of London Ser. A 60, 1897, 477–489.