Abstract

Machine Learning is a branch of Computer Science that is concerned with designing systems that can learn from the provided input. Supervised Machine Learning is where the system needs to be first trained using already classified training data as opposed to an unsupervised system where no such training is required. Supervised learning comprises of 2 training techniques. Linear Regression predicts a continuous valued output. Logistic Regression, more commonly known as Classification predicts a discrete valued output. It is the algorithm for identifying to which of a set of categories a new observation belongs. In this paper we aim to assess whether a lump in a breast could be malignant (cancerous) or benign (non-cancerous) by Classification. The 2 features under consideration are Clump Thickness and Marginal Adhesion. Clump Thickness helps us detect cancerous cells as they are often grouped in multilayers whereas benign cells tend to be grouped in monolayers. Normal cells tend to stick together but Cancerous cells tend to lose this ability. So loss of Marginal Adhesion is a sign of malignancy. With the help of the sigmoid function, we find the Cost function of our data and minimize the sum of the squared errors over the training set. Using Gradient Descent we find the global
minimum of our Cost function and then calculate the parameters that fit our data. Finally we estimate the probability of the patient’s tumor being malignant or benign based on the values of these 2 features and the parameters.

References

8. Duda Ro & Hard PE, Pattern Classification and Scene Analysis, New York.

Index Terms

Computer Science  Artificial Intelligence

Keywords

Keywords are your own designated keywords which can be used for easy location of the manuscript using any search engines.