Chapter 1
Classification Models in Machine Learning Techniques

Amit Majumder

National Institute of Technology, Jamshedpur, India

ABSTRACT

Classification is the process of identifying, understanding, and grouping objects and ideas into specified categories. These pre-categorized training datasets are used by machine learning techniques to classify datasets into relevant and acceptable categories. Using the incoming training data, machine learning classifiers assess the chance or probability that the incoming data will fall into one of the established categories. One of categorization's most prominent applications is used by the largest email service providers of today: classifying emails as "spam" or "non-spam." In essence, classification is a form of "pattern recognition." Following the application of classification algorithms to the training data, the same pattern (similar number sequences, words, or attitudes, etc.) is found in future data sets. Classification falls within the category of supervised learning in the context of machine learning.

1. WHAT IS CLASSIFICATION?

Classification is the process of identifying, understanding, and grouping objects and ideas into specified categories. These pre-categorized training datasets are used by machine learning techniques to classify datasets into relevant and acceptable categories.

Using the incoming training data, machine learning classifiers assess the chance or probability that the incoming data will fall into one of the established categories. One of categorization's most prominent applications is used by the largest email service providers of today: classifying emails as "spam" or "non-spam" (Dada et al., 2019). In essence, classification is a form of "pattern recognition." Following the application of classification algorithms to the training data, the same pattern (similar number sequences, words, or attitudes, etc.) is found in future data sets.

Classification falls within the category of supervised learning in the context of machine learning. While learning is being done under supervision, the data being fed to an algorithm or network for clas-

sification has already been labelled, with the key characteristics/attributes having already been divided into distinct groups.

The act of sending data into a neural network or any machine learning algorithms and allowing it to recognise patterns in the data is known as "training" a model. To build a model for classification task it needs input data or input example in the format of features and the tag or the class label of the example. The data in feature format is processed during the training phase to generate predicted output. Both the features and the labels of the training data are provided to the network or machine learning algorithm as input during training for a supervised classification type work, but only the features are fed as input to the network during testing.

During training it needs data from a number of examples. For example, to classify a person as male or female it needs to consider data (feature data or attribute data e.g., hair length, height etc. of a person) of some male persons and some female persons. Here each person is considered as an example. The set of example's data which is also called a dataset is split into training and testing sets, which are two distinct sets of inputs. Since the model has already discovered the patterns in this collection of data, testing the classifier on the same dataset that it was trained on would be extremely biased. Instead, the dataset is divided into two sets: a training set that the classifier uses to refine its predictions, and a testing set that the classifier has never seen before. The primary objective of a classification algorithm is to determine the category of a given dataset, and these algorithms are frequently applied to forecast the results for categorical data.

An illustration of supervised categorization is shown in Figure 1. Breast cancer patients' benign and malignant tumours are identified in the dataset. By fitting a straight line through the data, the supervised classification algorithm will try to divide tumours into two groups. Based on the straight-line categorization, subsequent data can subsequently be categorised as benign or malignant. Only two discrete outputs

Figure 1. Classification example
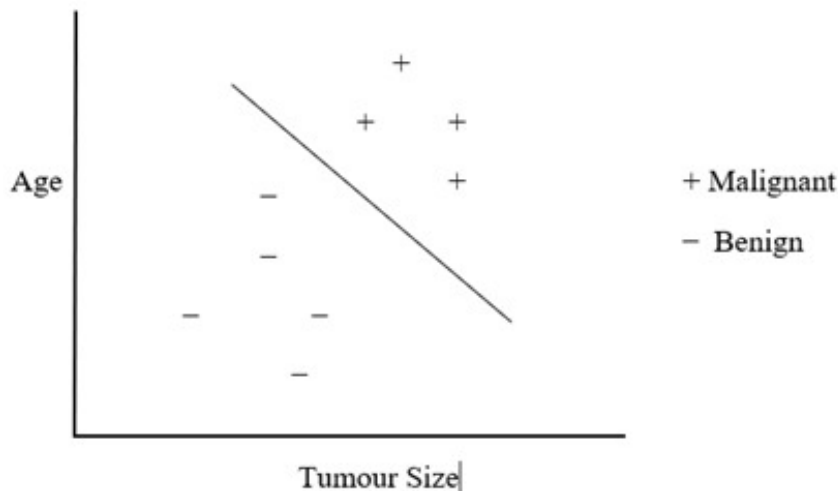
Related Content

Forecasting the Traits of Cyber Criminals Based on Case Studies
Nandini Bansod, Dinesh Baban Kamble, Rina Mishraand Megha Kuliha (2022). *Dark Web Pattern Recognition and Crime Analysis Using Machine Intelligence (pp. 220-234).*
www.irma-international.org/chapter/forecasting-the-traits-of-cyber-criminals-based-on-case-studies/304213

Performance Prediction and Optimization of Solar Water Heater via a Knowledge-Based Machine Learning Method
Hao Liand Zhijian Liu (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications (pp. 714-733).*
www.irma-international.org/chapter/performance-prediction-and-optimization-of-solar-water-heater-via-a-knowledge-based-machine-learning-method/307480

Object Detection in Fog Computing Using Machine Learning Algorithms
Peyakunta Bhargaviand Singaraju Jyothi (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications (pp. 472-485).*
www.irma-international.org/chapter/object-detection-in-fog-computing-using-machine-learning-algorithms/307468

Predicting Tomorrow's Air Quality Today: A Comprehensive Analysis of Machine Learning, Deep Learning, and Time Series Models in AQI Prediction
D. Narasimhanand M. Vanitha (2024). *Reshaping Environmental Science Through Machine Learning and IoT (pp. 311-323).*
www.irma-international.org/chapter/predicting-tomorrows-air-quality-today/346582

Deep Data Analytics: Future of Telehealth
Madhulika Bhatia, Madhurima Hoodaand Priya Gupta (2020). *Applications of Deep Learning and Big IoT on Personalized Healthcare Services (pp. 20-41).*
www.irma-international.org/chapter/deep-data-analytics/251232