4

SUPERVISED LEARNING

Said Jadid Abdulkadir

4.1 INTRODUCTION

ML is one of the AI's principal research fields. It includes developing, evaluating, implementing, and enforcing programs that are capable of learning [16]. According to Mitchell *et al.* [17], the definition of a general machine learning problem is stated as: "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E*". As stated by Mitchell and Mitchell [18], supervised learning, unsupervised learning, and reinforcement learning are three main categories of machine learning algorithms. This chapter focuses mainly on supervised learning where a model can learn from the inputs provided as a labelled dataset [19]. The meaning of a labelled dataset is that the desired output for each dataset is provided as well. The main purpose is to learn the function of mapping the input to the desired output. Common tasks of this type of learning are classification and regression.

In this chapter a clear description of supervised machine learning is explained, Figure 4.1 shows the main types of supervised learning approaches that can be applied to big data problems.

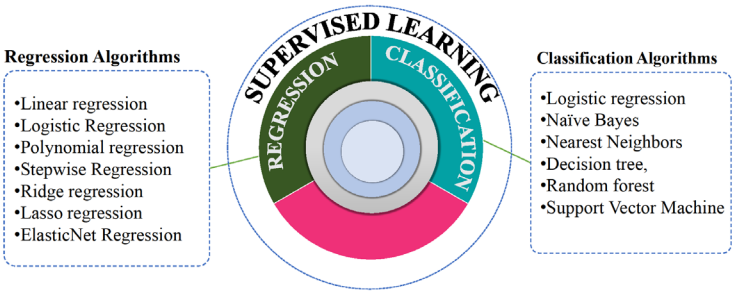

Figure 4.1 Supervised learning

In the supervised model, learning involves developing a method capable of learning utilising a training data set. This method can then be applied to unknown data to predict results. The aim is to design the feature in such a way that it performs well on all types of unknown input data. In two stages, the supervised learning process that consists of constructing and evaluating a mapping function is illustrated in Figure 4.1. The first step is to divide the dataset into two groups of samples: training dataset and testing dataset. The inputs represented by a test vector and at least one ideal output value are present in both training and test results. With the training data collection, train the mapping feature until it achieves an accurate mapping between the training data and the associated desired output. This happens for each training sample in the form of supervised learning, where using the cost function to find the error-(actual vs. expected output) to change the mapping function. The learned mapping feature is then tested by using test dataset. The test dataset contains data that is not included in the training phase and is an excellent indicator of the capability of the mapping algorithm in generalising to new data.

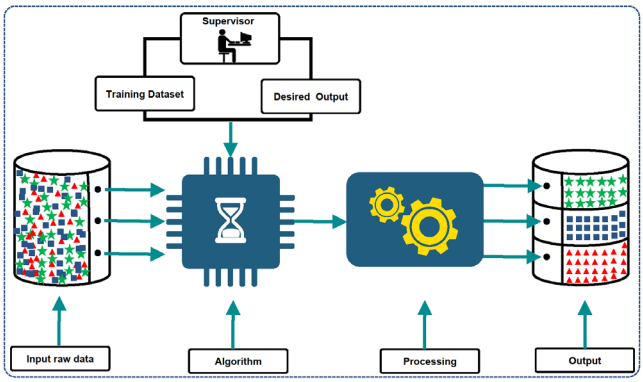

Figure 4.2 Supervised learning process

Common tasks of this type of learning are classification and regression [20]. In the following sections we will explain two types of supervised learning, regression, and classification.

4.2 REGRESSION

Regression learning is supervised learning that can be used to solve problems that have real or continuous values. Additionally, it is a prediction modelling approach that examines the relationship between a target (dependent) variable and (predictor) independent variable. This type of learning is utilised for modelling time series problems, forecasting, and determining the cause effect relation between variable [20]. Figure 4.3 shows an example of the learning regression.