T/UDOM/2017/02749

chapter 2

Supervised learning here the user provides the algorithm with pair of input and desired outputs and the algorithm finds away to produce the desired output. Supervised learning is divided into two types which are classification and regression problem .

Classification problem the aim is to predict a class label, which is a choice from a predefined list of possibilities. Classification is separated into binary and multiclass.

- In binary here we distinguish between exactly two classes, were one is positive and another is negative class .
- In multiclass here we distinguish more than two classes.

Regression :-this type does prediction from certain values. It predicts a continuous number or floating point. The difference between classification and regression is to look if there is a continuity in the values.

Generalization, overfitting , underfitting.

We say that the model may be able to generalize from the train set to test set if a model is able to make accurate predictions on unseen data. The aim is to build a model that is able to generalize as accurately as possible.

Underfitting is the process of choosing a too simple model. This model won't be able to provide a better prediction.

Where by for Overfittng we build a model which is too complex model for the amount of information.

Supervised learning Algorithm

- **k-Nearest Neighbors Algorithm-**is the simplest machine learning algorithm. When building the model consist only of storing training dataset. The algorithm finds the closest data points in the training dataset ts nearest neighbors. Were 'k' is th total number of neighbors chosen.

- **Decision Trees-**Decision trees are widely used models for classification and regression tasks. The algorithm starts from the root and then to the branches