NAME: DICKENS C ANTHONY

REG #: T/UDOM/2017/11340

PROG: BSC CS

Supervised learning

Are most useful and used whenever we want to predict a particular outcome from a given input, and we have examples of input/output pairs. But these input/output pairs are used to build machine learning models

There are two major types of supervised machine learning problems, called classification and regression,

i. **Classification**: Its goal refers to as prediction of a class label and it is divided into two parts namely binary and multiclass classifications

   i. binary: It distinguishes between exactly two cases
   ii. multiclass: classification between more than two classes Example, yes or no questions -spam/non spam

ii. **Regression:** It refers to as **a** prediction of a continuous number or a floating number in programming terms/real number in math.

   Examples of the predictions are,

- Prediction of a person's annual income from their education, age, and where they live.
- Predict the yield of a corn farm given attributes such as yields, weather, and number of employees working on the farm.

Generalization, overfitting and underfitting

If a model is able to make accurate predictions on unseen data, we say it is able to generalize from training set to the test set.

a. **Overfitting:** Refers to as a building the model that is too complex for the amount of data available. Overfitting occurs when you fit a model too closely to the training set but is not able to generalize to new data.

b. **undefitting;** Refers to as a choosing too simple a model when there is no possibility of capturing all the aspects of and variability in the data.

Model complexity and dataset

i. The larger variety of data points your dataset contains, the more complex a model you can use without overfitting.

ii. Larger datasets allow building more complex models.

iii. Hashing more data and building more complex models can often work wonders for supervised learning tasks.

Supervised learning algorithms k-

i. Nearest Neighbors

the model consists only of storing the training dataset. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset—its "nearest neighbors."

Other algorithms are as follows.

ii. Linear models for regression

iii. Linear regression

iv. Ridge regression

v. Lasso

vi. Linear models for classification

7) Logistic regression
8) Linear support vector machines
9) Linear models for multiclass classification
10) Naïve Bayes classifiers
11) Decision trees
12) Random fores

Reference

SCIENTISTS: Andreas C. Müller & Sarah Guido