BYTE QUEST

Vasavi College of Engineering

Department of Computer Science and Engineering

October 15, 2019

Volume 74

Contents:

Byte Quest is the article published by the CSE dept of Vasavi College of Engineering regarding the latest innovative Technologies and Software that have been emerged in the competitive world. The motto of this article is to update the people regarding the improvement in technology. The article is designed by the active participation of students under the guidance of faculty coordinators.

 Good, bad or indifferent if you are not investing in new technology, you are going to be left behind.

-Philip Green

 Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road.

-Stewart Brand

FACULTY CO-ORDINATORS

M.SUNDARI (ASST. PROFESSOR)

GARIMA JAIN (ASST. PROFESSOR)

STUDENT COORDINATORS

ESHWAR (4/4 CSE-A)

G CAROL (3/4 CSE-A)

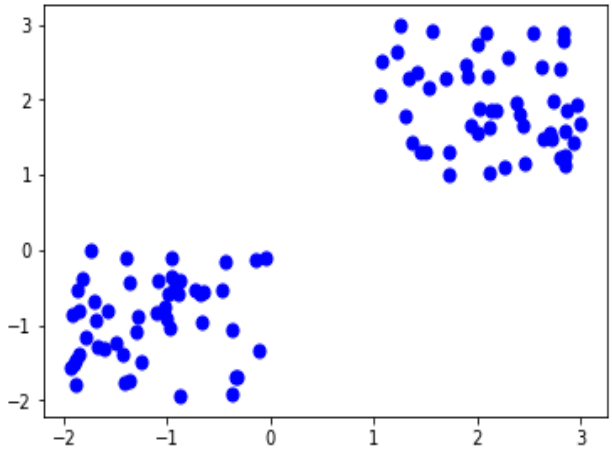
ABHINAV (2/4 CSE-A)

SREEEJA (4/4 CSE-B)

D. APARNA (3/4 CSE-B)

ANISHA (2/4 CSE-B)

K-MEANS CLUSTERING ALGORITHM

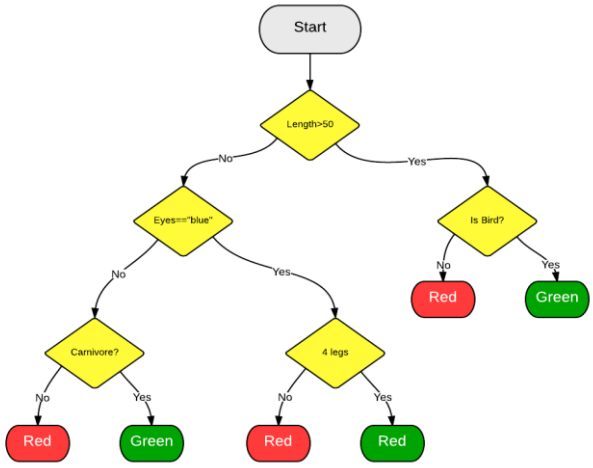


To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids

K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data. **k-means clustering** is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. Applying the 1-nearest neighbor classifier to the cluster centers obtained by *k*-means classifies new data into the existing clusters. This is known as nearest centroid classifier or Rocchio algorithm.

MANVITH REDDY (CSE-B 2/4)

RANDOM FORESTS ALGORITHM

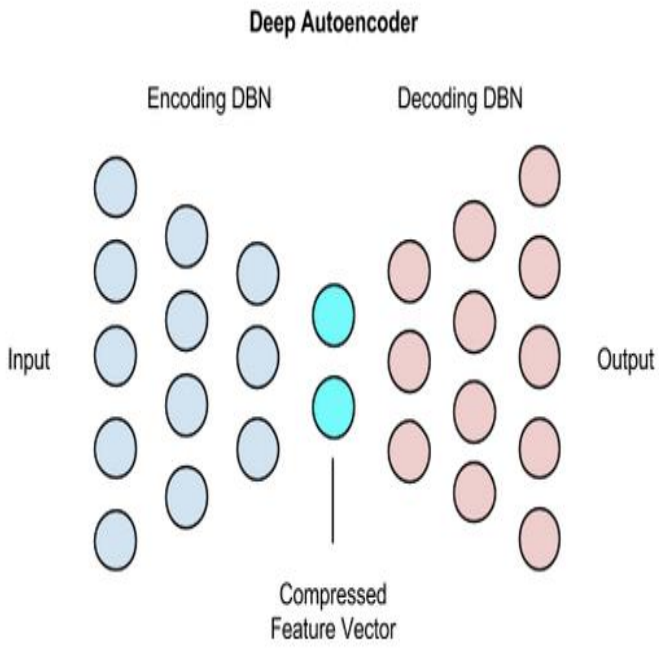


Random forests or **random decision forests** are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction

The early development of Breiman's notion of random forests was influenced by the work of Amit and Geman who introduced the idea of searching over a random subset of the available decisions when splitting a node, in the context of growing a single tree. The idea of random subspace selection from Ho was also influential in the design of random forests. In this method a forest of trees is grown, and variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each tree or each node. Finally, the idea of randomized node optimization, where the decision at each node is selected by a randomized procedure.

NAEHAL (CSE-B 2/4)

AUTOENCODERS



Autoencoders (AE) are a family of neural networks for which the input is the same as the output. They work by compressing the input into a latent-space representation and then reconstructing the output from this representation. One way to think of what deep learning does is as “A to B mappings,” says Andrew Ng, chief scientist at Baidu Research. “You can input an audio clip and output the transcript. That’s speech recognition.” As long as you have data to train the software, the possibilities are endless, he maintains. “You can input email, and the output could be: Is this spam or not?” Input loan applications, he says, and the output might be the likelihood a customer will repay it. Input usage patterns on a fleet of cars and the output could advise where to send a car next. With appropriate dimensionality and sparsity constraints, autoencoders can learn data projections that are more interesting than PCA or other basic techniques.

Rather making the facts complicated by having complex definitions, think of deep learning as a subset of a subset. Artificial Intelligence encircles a wide range of technologies and techniques that enable computers systems to unravel problems in ways that at least superficially resemble thinking. Within that sphere, there is that whole toolbox of enigmatic but important mathematical techniques which drives the motive of learning by experience. That subset is known to be machine learning. Finally, within machine learning is the smaller subcategory called deep learning (also known as deep structured learning or hierarchical learning)which is the application of artificial neural networks (ANNs) to learning tasks that contain more than one hidden layer.

Data-specific: Autoencoders are only able to compress data similar to what they have been trained on. An autoencoder which has been trained on human faces would not be performing well with images of modern buildings. This improvises the difference between autoencoders and MP3 kind of compression algorithms which only hold assumptions about sound in general, but not about specific types of sounds.

In the traditional architecture of autoencoders, it is not taken into account the fact that a signal can be seen as a sum of other signals. Convolutional Autoencoders (CAE), on the other way, use the convolution operator to accommodate this observation. Convolution operator allows filtering an input signal in order to extract some part of its content. They learn to encode the input in a set of simple signals and then try to reconstruct the input from them.

MOHAMMED AZAM (CSE-B 2/4)