Weekly report of lessons

Name: Ravi Pratap Singh
Roll No: 20CS60R60
The week : 11 Nov, 12 Nov, 13 Nov

The topics covered : Perceptron Modelling Neuron, Feedforward Network,Chain rule Computing gradient of neuron,Back Propagation ,Convergence of gradient descent,Clustering, K means Clustering, SSE ,Benefit & weakness of K Means , K-means++ ,hierarchical clustering, Clustering Affinity Search Technique(CAST), DBSCAN

Summary topic wise :

- The perceptron is a mathematical model of a biological neuron. We model this phenomenon in a perceptron by calculating the weighted sum of the inputs to represent the total strength of the input signals, and applying a step function on the sum to determine its output.
- Feedforward neural networks, or multilayer perceptrons, are the quintessential deep learning models. The goal of a feedforward network is to approximate some function f* .
- Since we have a neural network with just one layer and a loss function. That one layer is a simple fully-connected layer with only one neuron, numerous weights w_1, w_2, w_3…, a bias b, and a ReLU activation therefore we need a gradient of a neuron.According to chain rule the gradient will be : $\frac{\partial C}{\partial w^l_{jk}} = \frac{\partial C}{\partial z^l_j}\frac{\partial z^l_j}{\partial w^l_{jk}}$ by differentiation : $\frac{\partial C}{\partial w^l_{jk}} = \frac{\partial C}{\partial z^l_j} a^{l-1}{}_k$
- In fitting a neural network, backpropagation computes the gradient of the loss function with respect to the weights of the network for a single input–output example, and does so efficiently, unlike a naive direct computation of the gradient with respect to each weight individually.
- An iterative algorithm is said to converge to a solution if the value updates arrive at a fixed point Where the gradient is 0 and further updates do not change the estimate The algorithm may not actually converge , It may jitter around the local minimum and it may even diverge.
- Clustering is basically a type of unsupervised learning method .Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.
- The objective of K-means is to group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.
- SSE is the sum of the squared differences between each observation and its group's mean. It can be used as a measure of variation within a cluster. If all cases within a cluster are identical the SSE would then be equal to 0. $E = \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Benefit:K-means is guaranteed to converge at a quadratic rate. Linear time complexity in N,d and k . Weakness : Only detects well separated, compact hyperspherical clusters. Sensitive to noise and outlier points, may get stuck in local min.
- K-Means++ :The randomization of picking k-centroids points results in the problem of initialization sensitivity. The final formed clusters depend on how initial centroids were picked. K-Means++ is a smart centroid initialization technique and the rest of the algorithm is the same as that of K-Means. It chooses c1 randomly . The distance of x_i point from the farthest centroid can be computed by $d_i = max_{(j:1->m)} ||x_i - C_j||^2$.
- Hierarchical clustering is divided into two types **Agglomerative Hierarchical clustering Technique:** In this technique, initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed. In **Divisive Hierarchical clustering**, we consider all the data points as a single cluster and in each iteration, we separate the data points from the cluster which are not similar. In the end, we'll be left with n clusters. The Hierarchical clustering Technique can be visualized using a **Dendrogram.**
- The CAST has been found to be an attractive clustering procedure for module detection in gene expression data. CAST is a sequential procedure that defines clusters one at a time.
- The DBSCAN algorithm is used to find associations and structures in data that are hard to find manually.It is based on this intuitive notion of "clusters" and "noise". The key idea is that for each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points.

Any novel idea of yours out of the lessons : Making a Sum of Values for the row with numerical vs. row with more zeros can help model K means better. I Think this could be done with PCA .
Preparedness of the upcoming Quiz : Fair