**EXPERIMENT 4**

Rebecca Dias  182027 Roll no : 19 TE CMPN A2

Date of Submission:10-04-2021

**Aim**: - To implement Steepest Descent algorithm.

**Colab Link:-**

<https://colab.research.google.com/drive/11nXiWP6qCtUSTqepZ20Z3_rlgT4GWR3S?usp=sharing>

**Theory**:-

Mention what is optimization. What are its types and the differences between them?

Ans: Optimization is a process of obtaining the best results under any given circumstances and

choosing the best element from some set of available alternatives. Solving problems in which one

seeks to minimize or maximize a real function. There are perhaps hundreds of popular

optimization algorithms, and perhaps tens of algorithms to choose from in popular scientific code

libraries.

**Types:**

**Gradient Descent**

The gradient descent method is the most popular optimization method. The idea of this method is to update the variables iteratively in the (opposite) direction of the gradients of the objective function. With every update, this method guides the model to find the target and gradually converge to the optimal value of the objective function.

**Stochastic Gradient Descent**

Stochastic gradient descent (SGD) was proposed to address the computational complexity involved in each iteration for large scale data. Taking the values and adjusting them iteratively based on different parameters in order to reduce the loss function is called back-propagation. In this method, one sample randomly used to update the gradient(theta) per iteration instead of directly calculating the exact value of the gradient. The stochastic gradient is an unbiased estimate of the real gradient.

**Adaptive Learning Rate Method**

Learning rate is one of the key hyper parameters that undergo optimization. Learning rate decides whether the model will skip certain portions of the data. If the learning rate is high, then the model might miss on subtler aspects of the data. If it is low, then it is desirable for real-world applications. Learning rate has a great influence on SGD. Setting the right value of the learning rate can be challenging. Adaptive methods were proposed to this tuning automatically.

**Conjugate Gradient Method**

The conjugate gradient (CG) approach is used for solving large scale linear systems of equations and nonlinear optimization problems. The first-order methods have a slow convergence speed. Whereas, the second-order methods are resource-heavy. Conjugate gradient optimization is an intermediate algorithm, which combines the advantages of first-order information while ensuring the convergence speeds of high-order methods.

**Derivative-Free Optimization**

For some optimization problems, it can always be approached through a gradient because the derivative of the objective function may not exist or is not easy to calculate. This is where derivative-free optimization comes into the picture. It uses a heuristic algorithm that chooses methods that have already worked well, rather than derives solutions systematically.

**Part III**

Give a brief overview on different evolutionary techniques used in optimizations.

The genetic algorithm (GA) is one of the oldest and most known optimization techniques, which are based on nature. In the GA, the search for solution space imitates the natural process which takes place in the environment, and the Darwinian theory of species evolution is taken into consideration.

In GAs, we have a population of individuals; each, called a chromosome, represents a potential solution to the problem. The problem being solved is defined by the objective function. Depending on how “good” the given individual is fitted to the objective function, the value which represents its quality is attributed to it. This value is referred to as the fitness of the individual, and it is a main evaluating factor.

Highly valued individuals have a better chance to be selected to the new generation of the population.

In GAs, we have three operators:

**selection** (a new population of individuals is created based on the fitness values of individuals from the previous generation),

**crossover** (typically parts of individuals are exchanged between two individuals selected to the crossover), and

**mutation** (the values of particular genes are changed randomly).

**Conclusion:**

During this experiment , we implemented the Steepest Descent algorithm and learnt about the

PCA algorithm. Principal Component Analysis, or PCA, is a dimensionality-reduction method

that is often used to reduce the dimensionality of large data sets, by transforming a large set of

variables into a smaller one that still contains most of the information in the large set.

We, standardize the dataset then calculate the covariance matrix for the features in the dataset

and calculate the eigenvalues and eigenvectors for the covariance matrix.Sort eigenvalues and

their corresponding eigenvectors, pick k eigenvalues and form a matrix of eigenvectors and

finally transform the original matrix.