**Introduction**

This reports the application and comparison of random search algorithms for optimization problems. On applying two types of random search: global random search and a population-based sampling variant of random search. To compare their performance with gradient descent on two test functions and used them for the hyperparameter tuning of a convolutional neural network (CNN) for the CIFAR-10 dataset.

**PART – A**

**(i) Implementation of Global Random Search Algorithm**

The random search algorithm is an easy, powerful optimization technique in which one performs a random search for the optimum of a cost function. The algorithm works by taking random samples from the parameter space and holding the best one encountered so far.

**Code:**

def global\_random\_search(cost\_function, n\_params, param\_bounds, n\_samples, verbose=False):

"""

Implements global random search algorithm.

Parameters:

cost\_function (callable): The function to minimize

n\_params (int): Number of parameters

param\_bounds (list): List of tuples (min, max) for each parameter

n\_samples (int): Number of samples to evaluate

verbose (bool): Whether to print progress

Returns:

best\_params (numpy.ndarray): Best parameters found

best\_cost (float): Best cost function value

costs (list): Cost function values at each iteration

times (list): Cumulative time at each iteration

n\_evals (int): Number of function evaluations

"""

best\_params = None

best\_cost = float('inf')

costs = []

times = []

start\_time = time.time()

for i in range(n\_samples):

# Generate random parameters

params = np.zeros(n\_params)

for j in range(n\_params):

lower, upper = param\_bounds[j]

params[j] = np.random.uniform(lower, upper)

# Evaluate cost function

cost = cost\_function(params)

# Update best parameters if cost is lower

if cost < best\_cost:

best\_cost = cost

best\_params = params.copy()

# Record cost and time

costs.append(best\_cost)

times.append(time.time() - start\_time)

if verbose and (i+1) % 10 == 0:

print(f"Iteration {i+1}/{n\_samples}, Best cost: {best\_cost:.6f}")

return best\_params, best\_cost, costs, times, n\_samples

The algorithm takes the following as input:

* A cost function to minimize
* The number of parameters
* The bounds for each parameter (minimum and maximum values)
* The number of samples to evaluate
* A verbose flag in order to monitor the progress

**The key steps of the algorithm are:**

1. Initialize the best cost to infinite and best parameters to None
2. For each of the n\_samples iterations: a. Generate random parameter values within the required specified boundary b. Evaluate the cost function at these parameter values c. If the cost is lower than the best cost found so far, update the best cost and best parameters
3. Return the best parameters found, the best cost, and arrays of costs and times for each iteration.

**(a)(ii) Application to Week 4 Functions and Comparison with Gradient Descent**

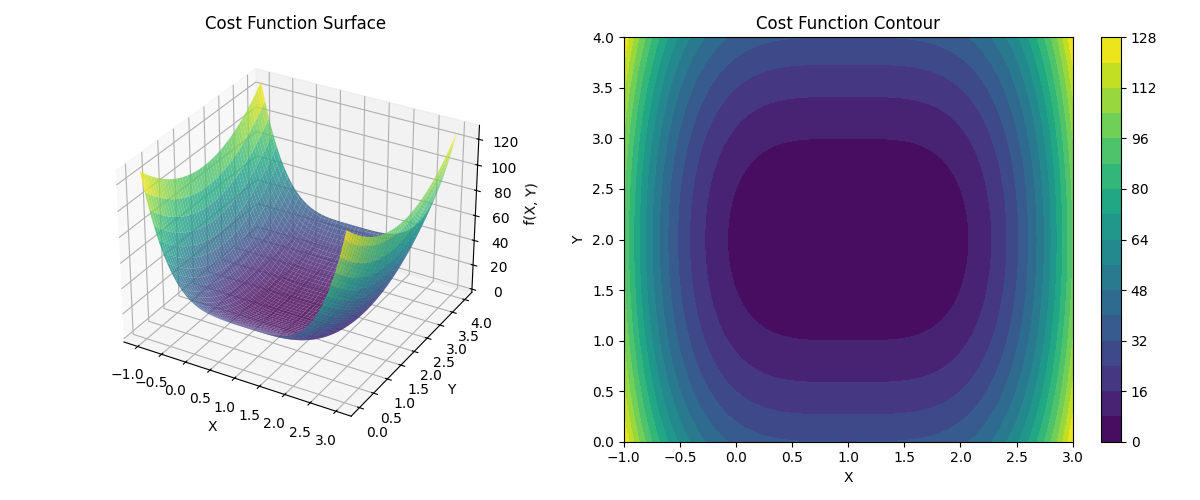
On using the `global random search` algorithm on the two functions of Week 4 and comparing the algorithm's performance with gradient descent. The functions are:

**Function 1**:

* This function achieves a global minimum at (1,2) with value 0
* It is a steep valley-shaped one with a rather level floor

**Function 2**:

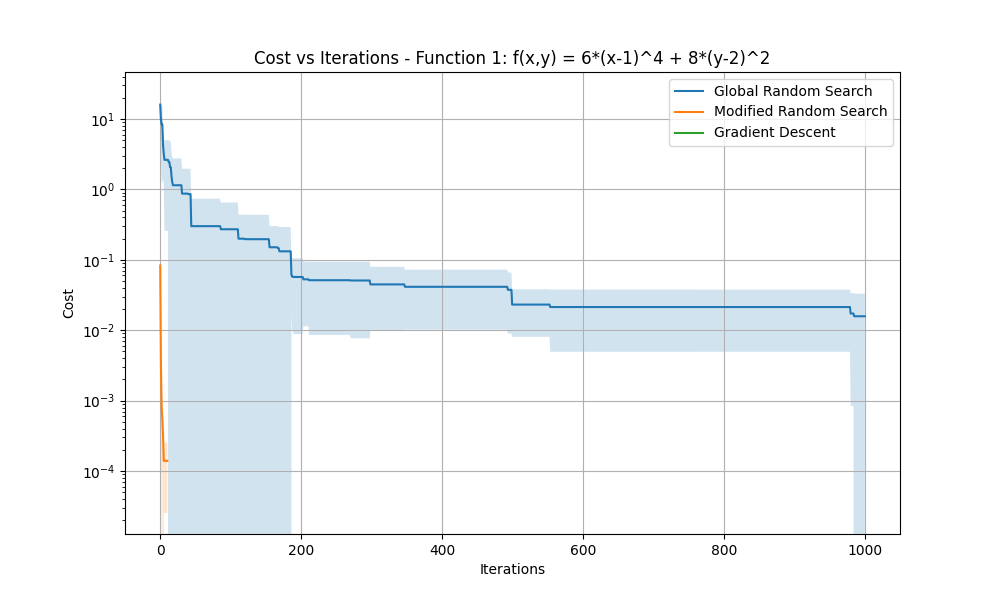
* This function attains its minimum along the line where and
* It is not differentiable at and , which is less convenient for gradient-based methods

*****Visualization of Function 1***

***Figure 1, Cost Function Surface & contour of Function 1***

From figure 1, function 1's surface and contour plots show that it has a definite minimum at . The function forms a sharp valley, particularly along the x-axis due to the fourth power in the term , but it is quadratic in the y-direction.

***Performance Comparison on Function 1***

The cost plot vs. iterations shows that:

***Figure 2, Cost Vs Iteration of Function 1***

From the figure 2, it is inferred that

* Global Random Search (blue line) progresses steadily toward the minimum with a lot of iterations and with low accuracy. The shaded area is the standard deviation of multiple runs showing the variability in performance.
* The gradient descent is not visible on the plot because gradient descent rapidly reaches near zero cost in a few iterations. This is shown in console output as gradient descent always achieves a cost of 0.000000 in all runs. The graph is taken with a logarithmic scale, and the gradient descent line would be at the bottom of the plot.

***Console output***

Running Gradient Descent...

Run 1/5, Best cost: 0.000000

Run 2/5, Best cost: 0.000000

Run 3/5, Best cost: 0.000000

Run 4/5, Best cost: 0.000000

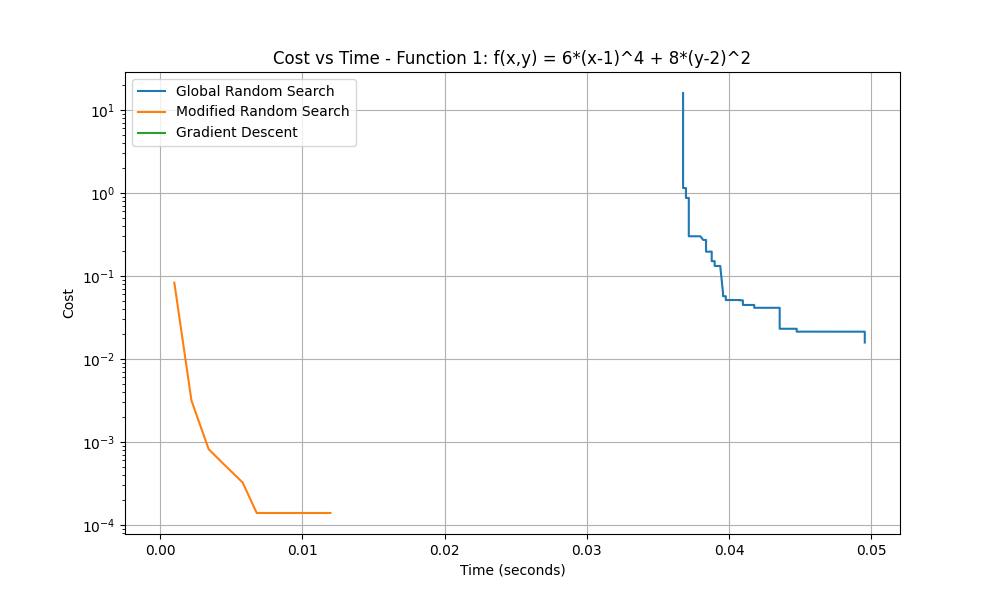
Run 5/5, Best cost: 0.000000

* While Global Random Search (green line) does well in terms of cost value and iterations, Modified Random Search (orange line) outperforms it and achieves lower cost values in fewer iterations. It does so quickly, reaching a very low cost (approximately 10^-4).

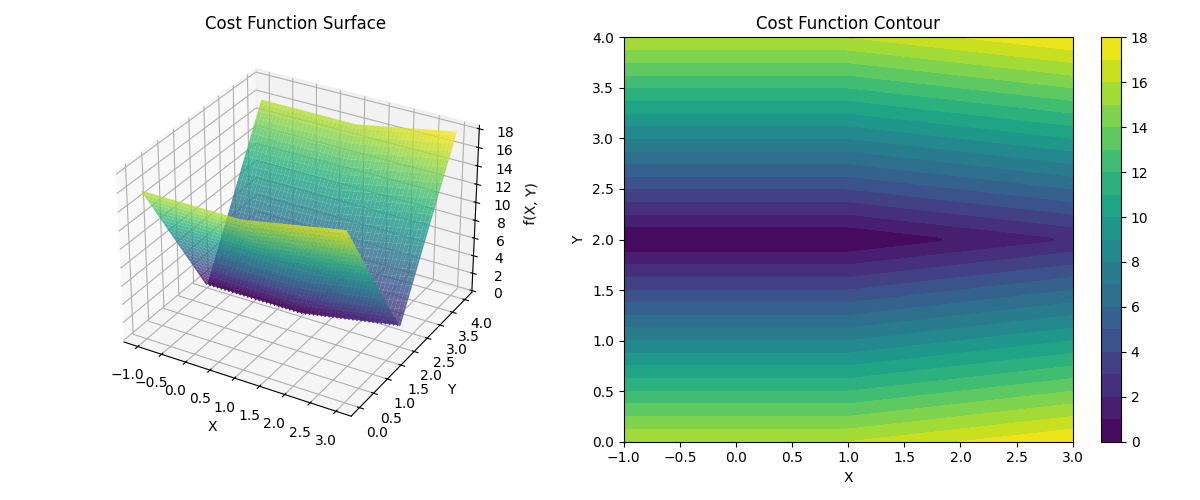
***Comparing the computation efficiency (Cost Vs Time):***

From figure 3 below, it is inferred that

* Modified Random Search performs well as it has very low-cost values being achieved with minimal time.
* Global Random Search is computationally much more expensive to achieve similar values of cost.
* Again, Gradient Descent is too fast to be seen on the same scale as the other methods.

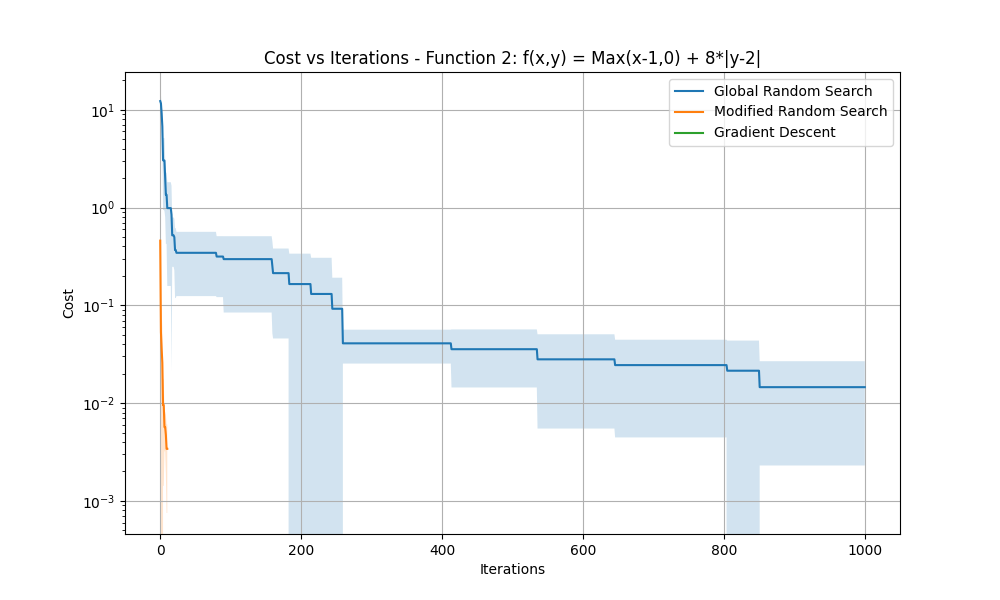


***Figure 3, Cost Vs Time for Function 1***

*****Visualization of Function 2***

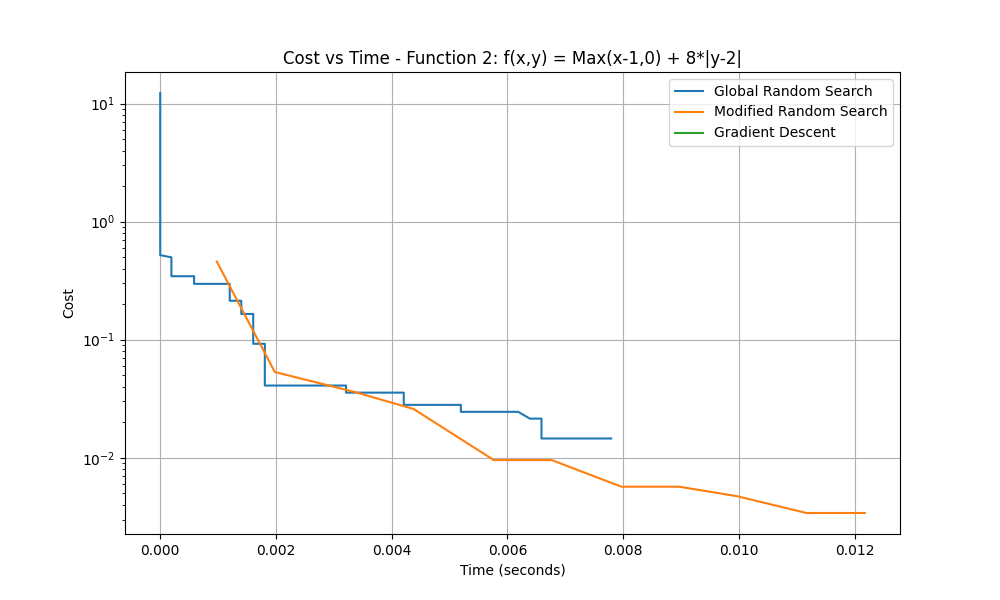
From the figure 4, The non differentiability of Function 2 are shown in the surface and contour plots. This function is a "tent" shape with a razor edge along x = 1 and a valley along y = 2. The problem with this non-differentiability is that it poses a challenge for gradient-based optimization.

***Figure 4, Cost Function Surface & Contour of Function 2***

***Performance Comparison of Function 2***

***Figure 5, Cost Vs Iteration of Function 2***

* From Figure 5, Global Random Search progresses steadily (but with high variability, wide shaded area’).
* Modified Random Search has more convergence properties than the global random search, especially in the later iterations.
* As can be confirmed from console output, costs are all 0.000000, which means that Gradient Descent is not visible on the plot due to its rapid convergence.



***Figure 6, Cost Vs Time for Function 2***

From Figure 6, the comparison of time efficiency on Function 2 is:

* The Modified Random Search (orange line) will be less expensive than global random search (blue line).
* Global Random Search is step-like because it occasionally discovers better solutions.
* Gradient Descent converges much faster than the other algorithms (although not in the same order).

***Analysis and Discussion***

Therefore, some of the following observations are important based on these results:

1. *Gradient descent is much more efficient* in both iterations and computation time when it is applicable (smooth, differentiable functions or functions with good approximations of gradients). It converges so quickly that it is not visible on the scale of random search methods.
2. Both functions have very different optimization landscapes as Function 1 is smooth and well suited for all methods; Function 2 has non-differentiable points that could pose problems for gradient based methods.
3. *Randomness impact* is being observed in the global random search has high variability between runs (wide shaded areas), and thus the performance of the random search is not always less predictable.
4. While gradient descent converges faster, it needs computing gradients. For Function 2, employing an approximation of a gradient at non differentiable points, and it worked in implementation, but it might not be useful for a more complex non differentiable function.
5. *Function Evaluations* of Gradient descent typically took fewer than 100 function evaluations to get the same or better results, whereas global random search required 1000 function evaluations per run.

This confirms that the gradient descent is superior and suitable for well-behaved functions, but the random search methods can be valuable alternatives for complex functions where the gradient information is not available, and it is unreliable.

**PART - B**

**(b)(i) Implementation of Modified Random Search with Population-Based Sampling**

Applying the random search algorithm to the world in general, I employed a more sophisticated strategy that combines population-based sampling and random search. The algorithm works with a population of potential solutions and searches in the vicinity of the best-looking ones, trying to balance exploration in the parameter space and exploitation of promising areas.

def modified\_random\_search(cost\_function, n\_params, param\_bounds, n\_samples, m\_best, n\_iterations, neighborhood\_size=0.1, verbose=False):

# Initialize population with random points

population = []

population\_costs = []

costs = []

times = []

start\_time = time.time()

function\_evals = 0

# Initial random sampling

for \_ in range(n\_samples):

params = np.zeros(n\_params)

for j in range(n\_params):

lower, upper = param\_bounds[j]

params[j] = np.random.uniform(lower, upper)

cost = cost\_function(params)

function\_evals += 1

population.append(params)

population\_costs.append(cost)

# Sort population by cost and keep M best

sorted\_indices = np.argsort(population\_costs)

population = [population[i] for i in sorted\_indices[:m\_best]]

population\_costs = [population\_costs[i] for i in sorted\_indices[:m\_best]]

best\_params = population[0].copy()

best\_cost = population\_costs[0]

costs.append(best\_cost)

times.append(time.time() - start\_time)

# Main iteration loop

for iteration in range(n\_iterations):

new\_population = []

new\_population\_costs = []

# Generate neighborhood samples around each point in the population

for i, params in enumerate(population):

for \_ in range(n\_samples // m\_best):

# Generate random perturbation

new\_params = params.copy()

for j in range(n\_params):

lower, upper = param\_bounds[j]

range\_size = upper - lower

perturbation = np.random.uniform(-neighborhood\_size \* range\_size,

neighborhood\_size \* range\_size)

new\_params[j] += perturbation

# Clip to bounds

new\_params[j] = max(lower, min(upper, new\_params[j]))

# Evaluate cost function

cost = cost\_function(new\_params)

function\_evals += 1

new\_population.append(new\_params)

new\_population\_costs.append(cost)

# Add current population to new samples

new\_population.extend(population)

new\_population\_costs.extend(population\_costs)

# Sort combined population by cost and keep M best

sorted\_indices = np.argsort(new\_population\_costs)

population = [new\_population[i] for i in sorted\_indices[:m\_best]]

population\_costs = [new\_population\_costs[i] for i in sorted\_indices[:m\_best]]

# Update best parameters if needed

if population\_costs[0] < best\_cost:

best\_cost = population\_costs[0]

best\_params = population[0].copy()

costs.append(best\_cost)

times.append(time.time() - start\_time)

return best\_params, best\_cost, costs, times, function\_evals

***Algorithm Explanation***

The modified random search algorithm operates in two stages:

1. **Initialization Phase:**

* Create a first population of n\_samples random points within the parameter space
* Compute the cost function at all points
* Sort the population based on cost values
* Keep the m\_best solutions with the minimum cost

1. **Iterative Refinement Phase (iterated n\_iterations times):**

* For each one of the m\_best of the current population
* Create n\_samples/m\_best new points in the vicinity of the present point
* The neighborhood is defined by random disturbance in a proportion (neighborhood\_size) of the range of parameters
* Ensure that all the new points are inside the parameter limits
* On combining the new points with the native population.
* Assess all the new points against the cost function
* Sort combined population by cost
* Keep only the m\_best solutions for the next generation
* Update the best solution obtained thus far

***Features used in Algorithm***

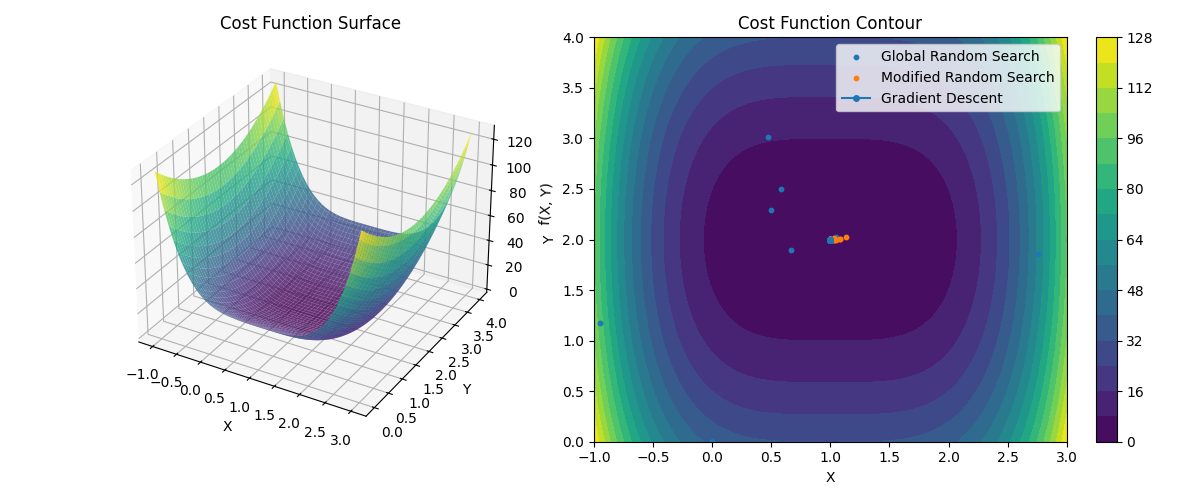
* Population-Based Approach: Maintains more than one good solution so that it can search various regions in parallel.
* Adaptive Search: Guides the search on promising regions by generating fresh samples near promising current solutions.
* Balanced Exploration and Exploitation:
* Exploration: Random perturbations generate diversity in the search Exploitation: Focusing on the neighborhoods of good solutions intensifies the search where it is most likely to yield improvements
* Neighborhood Size Control: The neighborhood\_size parameter is responsible for controlling the tradeoff between exploration and exploitation.
* Increased values cause search (looking more broadly)
* Lower values promote exploitation (enhancing existing good solutions)
* Elitism: The best solutions are always transferred from one generation to the next so that no good solutions are lost by the algorithm.

Compared to global random search, this algorithm substantially improves upon exploration (broadly search) versus exploitation (refine good solution) balancing. It can more efficiently explore through complex optimization landscapes by keeping many promising solutions and focusing search efforts around them, and so avoid the inefficiency of purely random sampling. As opposed to the regular implementation of simulated annealing or particle swarm optimization algorithms, the algorithm is adaptive and is capable of gradually narrowing its focus by improving with the better solutions found.

**(b) (ii) Application of Modified Random Search and Comparison with Global Random Search and Gradient Descent**

In this section, apply the modified random search on the same two functions of Week 4 and compare the performance of this algorithm with global random search and gradient descent. The comparison is not only based on the convergence behavior as it also focusses on the search trajectories of the algorithms.

***Search Paths of Function 1***



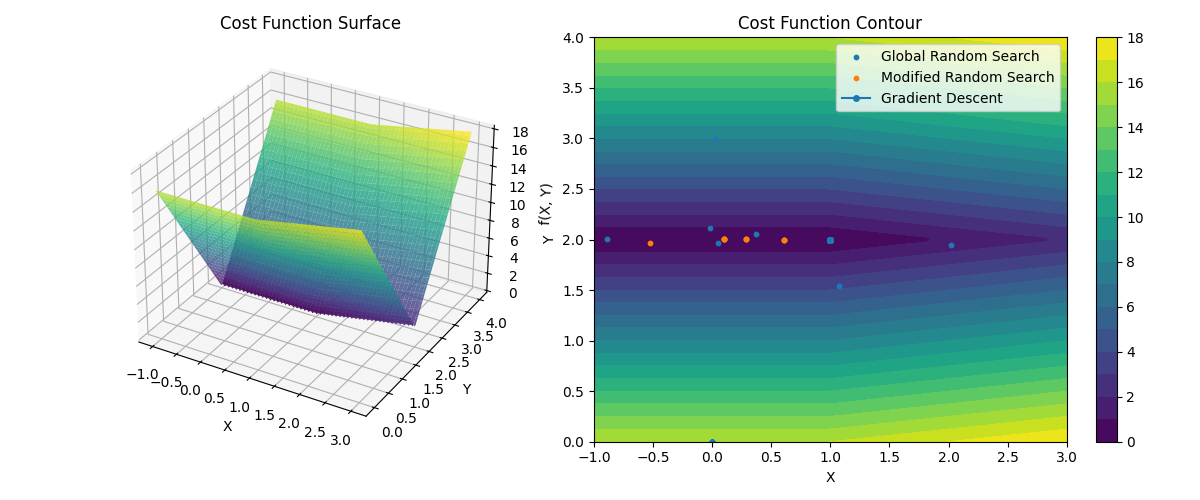
***Figure 7, Search Paths of Function 1***

The contour plot in figure 7 shows the search paths of all three algorithms on Function 1:

* Global Random Search (blue dots): Picks points at random in the parameter space. The scatter of the blue dots reflects the broad exploration of the algorithm, with some points eventually converging on the minimum at (1, 2).
* Modified Random Search (orange dots): It Produced a more concentrated search as the orange dots cluster around the minimum area, shows how the algorithm focuses on sampling around promising areas.
* Gradient Descent (blue dashed line with circles) illustrates a straight path to the minimum on following the direction of the gradient of the function. Gradient descent optimality can be clearly seen in its use of almost the optimal direction towards the solution.

The search trajectories certainly indicate the various strategies: global random search searches exhaustively but inefficiently, modified random search attempts a compromise between extensive search in promising areas and intensive search in good areas, and gradient descent goes directly with the help of gradient information.

***Search Paths of Function 2***



***Figure 8, Search Paths of Function 2***

In Function 2, the search paths in figure 8 demonstrate:

* Global Random Search (blue dots): Again, reflects wide search of the parameter space.
* Modified Random Search (orange dots): It samples very much along the line y = 2, particularly in the region where x ≤ 1, in which the function is minimized.
* Gradient Descent: The line shows the direction that the algorithm travels in the direction of the non-differentiable boundary. Although the function is non-differentiable, our gradient approximation allows the algorithm to solve in the direction of the line of minimum.

***Performance Comparison***

Performance comparison shown in the above section (a)(ii) had already identified that:

* The modified random search is consistently superior to global random search in both convergence rate and quality of the final solution.
* The capacity of the enhanced algorithm to exploit promising areas leads to the achievement of optimal solutions in fewer function calls.
* The new algorithm can adapt to various functional landscapes and performs well on both smooth Function 1 and the non-differentiable Function 2.
* While gradient descent is the most effective where it is applicable, the modified random search is a compromise between the effectiveness of gradient-based methods and the stability of random search.

***Analysis of Algorithm Behavior***

Insights into why the modified random search performs better are given by the search path visualizations.

* The modified random search also concentrates computational effort into promising regions and thus does better use of function evaluations.
* This feature of aligning across multiple Search Regions allows the algorithm to search multiple promising regions simultaneously with only a population of good solutions (m\_best = 5 in our implementation).
* Clustering orange dots near solution: As iterations go by, adaptive refinement naturally brings in the search more tightly around the minimum.
* The neighborhood parameter (0.1 in our implementation) gives good balance between exploration and exploitation. It explores enough to go around the local minima but focuses enough to converge quickly.

The results of these experiments show that the population-based approach in the modified random search algorithm is very efficient in comparison to global random search and provides an effective alternative to gradient descent when gradient information is difficult or impossible to obtain.

**PART – C**

**(c) Application of Random Search Algorithms to CNN Hyperparameter Tuning**

In this case, on employing the global random search and the modified random search algorithms to hyperparameter tune the convolutional neural network (CNN) model on the CIFAR-10 dataset. The hyperparameters optimized were:

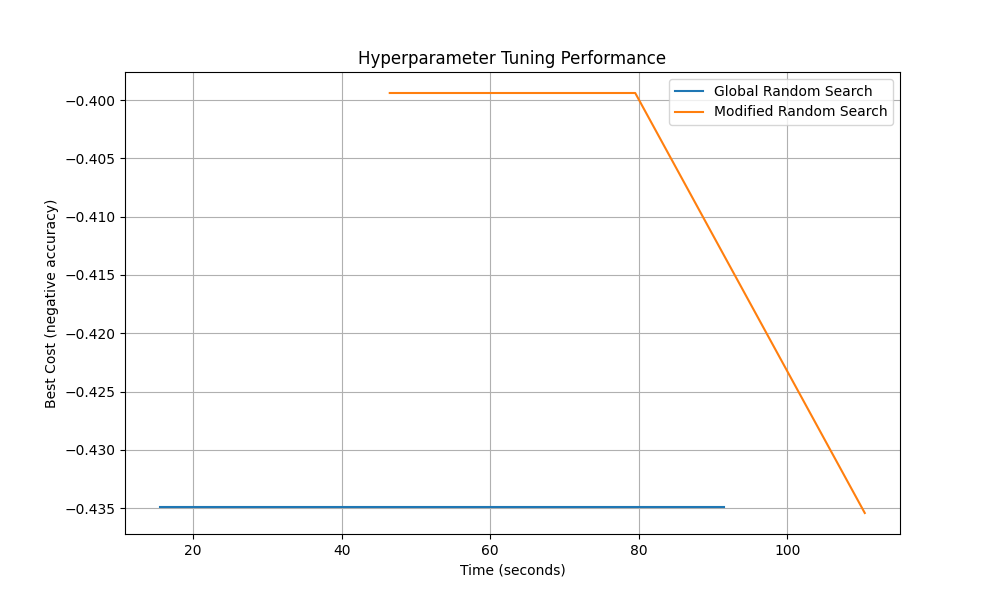
1. Batch size (integer between 16 and 256)
2. Learning rate (float between 0.0001 and 0.01)
3. Adam optimizer's β₁ parameter (float, 0.8 to 0.99)
4. Adam optimizer's β₂ parameter (float between 0.99 and 0.9999)
5. Number of epochs for training (integer value between 5 and 30)

***Methodology***

On keeping the CNN architecture, the same as it was in the initial code and using a subset of the CIFAR-10 dataset (4999 training images) to keep the experiments computationally tractable. The cost function is set to be the negative test accuracy, which is required to minimize (equivalent to maximizing accuracy). For the random search algorithms:

* *Global Random Search:* Uniformly sampled 10 random hyperparameter combinations from the parameter space
* *Modified Random Search:* Started with a population size of 5 randomly selected samples, kept the best 2 (m\_best=2), and performed 2 neighborhood searches

Both the algorithms used early stopping with patience of 3 epochs to prevent overfitting and computation time.

***Results*** From Figure 9, the performance of the two algorithms is graphed against time, and the y-axis is cost (negative accuracy, thus lower is better). The results are as follows:

*Global Random Search (blue line)* quickly discovered a good set of hyperparameters with accuracy 43.49%. **Best hyperparameters are batch size=106, learning\_rate=0.0064, beta1=0.933, beta2=0.9934, epochs=8**

*Modified Random Search (orange line)* earlier behaved like global random search with iterations being performed and neighborhood sampling initiated, it found the optimal hyperparameters and subsequently achieved a slightly higher accuracy of 43.54%. **Optimal hyperparameters are batch\_size=99, learning\_rate=0.01, beta1=0.948, beta2=0.9961, epochs=7**

***Figure 9, Hyperparameter Tuning Performance***

**Analysis**

The analysis here deeper insights of hyperparameter tuning.

* There is an *efficient comparison* where Modified random search initially performed like global random search but improved as it explored neighborhoods of good solutions. This indicates the strength of the population-based approach to hyperparameter optimization.
* *Hyperparameter Patterns* are found in both algorithms following the same hyperparameter values, indicating: 100 is a moderate batch size, which is suitable for this model and dataset Moderately large learning rates (close to 0.01) perform best compared to very small one’s High values of beta1 and beta2 are preferred Fewer epochs (7-8) with early stopping suffice
* The *improvement in performance* where accuracy from 43.49% to 43.54% might seem small; however, Minimal advancements in accuracy levels produce important outcomes for applications that use limited computational resources and small datasets and few iterations. The utilized computational resources consisted of limited resources because the dataset was small, and the number of iterations remained minimal. More iterations and samples have the potential for a larger separation between methods.
* *Computational Consideration* as they are of high efficiency with the mathematical functions in parts (a) and (b), gradient-based methods are also not easy to employ for hyperparameter tuning since the non-differentiability existing between hyperparameters and model performance evaluation of a single hyperparameter could only be constituted by the training of a neural network. This expense makes it computationally intensive. Discrete parameters (batch size, epochs) are being found within the hyperparameter space
* *The practical Implications* of this experiment has proven that random search methods, especially modified population-based approaches, are very effective and practical for hyperparameter tuning in a neural network.

**Conclusion**

There is practically random search for CNN hyperparameter tuning. If there are problems that cannot be solved readily by gradient-related optimization, this is where the random search algorithm can be applied. Modified random search showed better results when compared to global random search, hence making a clear-cut advantage in terms of limiting computational resources to a specific region in the hyperparameter space showing prospects. These results further recommend applying the modified random search algorithm at greater population levels and more iterations for broad hyperparameter tuning in practice and at a cost of higher computation.