

Replication Instructions

1 Environment Setup

1. Create a clean environment:

```
python -m venv bayesian_env
source bayesian_env/bin/activate # Linux/MacOS
# OR
bayesian_env\Scripts\activate # Windows
```

2. Install requirements:

```
pip install -r requirements.txt
```

2 Dataset Preparation

1. Make sure to have the original **datasets** from lab3 for baseline:

```
mkdir -p datasets
```

2. Download the benchmark datasets:

```
wget https://github.com/ideas-labo/ISE/tree/main/lab3/datasets
```

The datasets includes: 7z, Apache, Brotli, LLVM, PostgreSQL, Spear, Storm, and x264.

3 Reproducing Main Results

1. Run the baseline Random Search algorithm:

```
python main.py
```

This generates results in the **search_results** directory.

2. Run Bayesian Optimization method:

```
python bayesian_optimization.py
```

This generates results in the **bo_search_results** directory.

3. Run the comparison framework with 5 repetitions:

```
python comparison_methods.py --budget 100 --repetitions 5 --output_dir
comparison
```

This generates a comprehensive comparison in the **comparison** directory.

4 Reproducing Visualizations

1. Generate the convergence plot for Apache (Figure 1):

```
python comparison_convergence.py
```

2. Generate the performance improvement plot (Figure 2):

```
python comparison_improvement.py
```

3. Generate the search efficiency analysis (Figure 3):

```
python comparison_efficiency.py
```

If the results differ slightly, this is normal due to the randomness in both algorithms. However, the overall trends and relative improvements should be consistent.