**YAHAYA MISTURA**

**20CF027312**

**CEN524 REPORT**

**1. Findings**

**Synthetic Data (50 samples):**

After normalization of "hours studied," gradient descent was solved in 100 iterations to

1. Final MSE ≈ 0.48
2. Final MAE ≈ 0.54
3. Trials on three hand-tuned sets of parameters showed MSE reducing from ~10.2 to ~5.1 as (w,b) approached the actual values, with MAE reducing from ~2.8 to ~1.9.
4. 0.01 learning rate traded off between speed and stability: high rates overshot (diverged), low rates took >500 iterations to reach similar error.

**Real‑World Data (California Housing, MedInc feature):**

1. Gradient descent on the single normalized feature took converging longer, with final MSE ≈ 0.85 and MAE ≈ 0.72.
2. Using scikit‑learn's `LinearRegression` yielded a closed‑form solution in a single step (no iterations) with MSE ≈ 0.82 and MAE ≈ 0.70.

**2. Challenges Faced**

1. **Broadcasting Errors:** My initial synthetic‑target creation mixed 2D and 1D arrays, yielding a 50×50 matrix. I fixed it by flattening `X` and `y' to 1D before gradient descent.
2. **Learning Rate Tuning**: Choosing a sensible η required trial: 0.1 diverged, 0.001 converged too slowly. I selected 0.01 after the MSE plot comparison.
3. **Noisy vs. Clean Data:** Noisy real data contained outliers and noise, leading to slower convergence and higher final error than the synthetic, clean scenario.

**3. Suggested Improvement**

Employ **mini-batch gradient descent** with adaptive learning rates (e.g., Adam). This would:

1. Stabilize convergence on noisy real data
2. Reduce sensitivity to hyperparameters
3. Accelerate training with vectorized batch updates

**4. Answers to Questions**

**How does normalization alter the feature values?**

It normalizes features to unit variance and brings them to mean zero such that no individual feature dominates the gradient and converges quickly.

**Why is MSE more sensitive to significant errors than MAE?**

MSE squares up errors, and significant errors contribute more, so it is outlier-sensitive; MAE treats all errors proportionately.

**How does learning rate affect convergence?**

Slow progress with a low rate (e.g., 0.001); over­shoot of minima and possible divergence at a high rate (e.g., 0.1).

**Why might the model perform differently on real vs. synthetic data?**

Synthetic data models an identifiable linear rule with adjustable noise; real data has noise, latent structure, and non‑linear interactions.

MSE gives a smooth, differentiable surface but is outlier sensitive; MAE is robust but has a non‑differentiable point at zero error.

**What challenges arise when scaling to multiple features?**

Without normalization, differently scaled features disrupt convergence; increasing parameters results in a risk of multicollinearity and requires careful tuning of hyperparameters.

**How does gradient descent differ from scikit‑learn's `LinearRegression`?**

Gradient descent is tunable and iterative but generally slower; `LinearRegression` uses an analytic or optimized solver for a single-step, direct solution.