a. Based on accuracy, which dimensionality reduction method worked the best?

The **Simulated Annealing** method (Part 3) achieved the highest accuracy of **97.33%**, making it the best-performing feature selection technique.

- PCA (Part 2): 91.33%
- Genetic Algorithm (Part 4): 96.00%
- Baseline using original iris features (Part 1): 90.67%

b. For each of the two other methods, explain why you think they did not perform as well as the best one.

- **PCA:**
  PCA is an **unsupervised** technique: it seeks to maximize variance across features, not classification separability. The first principal component (z1) explained over 92% of variance, but that variance did not align perfectly with the features most relevant for distinguishing Iris classes. As a result, PCA alone underperformed compared to supervised search methods.

- **Genetic Algorithm:**
  The GA performed well but slightly below Simulated Annealing. Its accuracy plateaued early (96%) because the population quickly converged on a simpler subset (`['petal_width']`), reducing exploration diversity. With a small population and limited mutation, it likely got trapped in a local optimum.

c. Did the best dimensionality reduction method produce a better accuracy than using none (i.e., the results of Part 1)? Explain possible reasons.

Yes.
Simulated Annealing (97.33%) outperformed the baseline Decision Tree (90.67%).
This improvement happened because the search algorithm **selected a more informative and compact subset** of features (`['sepal_length', 'petal_length', 'petal_width', 'z3']`), removing slightly redundant dimensions (such as sepal width) that added noise to the model.
The baseline used all features equally, while SA optimized feature choice for classification accuracy.

d. Did Part 2 (PCA) produce the same set of best features as Part 3 (Simulated Annealing)? Explain.

No.

- **PCA** selected only **z1**, a single transformed feature derived from all four inputs.
- **Simulated Annealing** selected a mixed subset: **`['sepal_length', 'petal_length', 'petal_width', 'z3']`**.

This difference is due to their objectives:

- PCA maximizes **variance** (unsupervised),
- Simulated Annealing maximizes **accuracy** (supervised).

Thus, PCA focuses on overall data spread, while SA focuses directly on predictive performance.

e. Did Part 2 (PCA) produce the same set of best features as Part 4 (Genetic Algorithm)? Explain.

No.

- **PCA** used **z1**,
- **GA** converged on **`['petal_width']`**.

PCA transforms the entire feature space into linear combinations, while GA searches discrete subsets of the original + PCA features. Their optimization goals differ, leading to different results. In fact, PCA aims for compact variance representation, GA for maximum supervised accuracy.

f. Did Part 3 (Simulated Annealing) produce the same set of best features as Part 4 (Genetic Algorithm)? Explain.

No.

- **Simulated Annealing:** `['sepal_length', 'petal_length', 'petal_width', 'z3']`
- **Genetic Algorithm:** `['petal_width']`

This difference arises from how each metaheuristic explores the search space:

- **Simulated Annealing** accepts occasional worse moves to escape local minima, exploring more broadly.
- **GA** depends on crossover and mutation but can converge prematurely if population diversity is low.

SA's probabilistic exploration allowed it to find a better-performing feature subset, while GA favored simplicity at the cost of slightly lower accuracy.