

**Question 4: For Part 2, using the PoV formula and the values from the eigenvalue matrix, show that the program calculated the PoV correctly.**

- So our Eigen Value output is the following:

```
[4.22484077 0.24224357 0.07852391 0.02368303]
```

to calculate, we can simply do

```
4.22484077
```

```
4.22484077 + 0.24224357 + 0.07852391 + 0.02368303
```

this leads to a correct accuracy of .926666 . Which is the same as what the program outputted.

**Question 5a: Based on accuracy which dimensionality reduction method, PCA, simulate annealing, or the genetic algorithm worked the best?**

- Based off of our results and confusion matrix from our analysis, we can say that the Genetic Algorithm works the best. We can make this conclusion because the accuracy of the genetic algorithm is higher than the PCA and simulated annealing method. Though, there can be some randomness, sometimes others did better than others, but **\*\* ON AVERAGE \*\***, the Genetic Algorithm will work best.

**Question 5b: For each of the two other methods, explain why you think it did not perform as well as the best one**

- So, to start with the PCA, it performs all operations off of one set of data, so it doesn't do anything with comparisons in the accuracy or feature categories. PCA also assumes that the features are linearly correlated, while GA doesn't as it doesn't make that assumption. Same thing with simulated annealing. Since we are doing random probability to get the best accuracy, we won't run into errors with linearity with our data.
- To go to the simulated Annealing portion, we can also look at the randomness factor to negatively correlate the accuracy of our model while computing. This means it can find a sub-optimum solution, but GA will iterate almost every possible set and get the best accuracy that way.

**Question 5c: Did the best dimensionality reduction method produce a better accuracy than using none (i.e. the results of Part 1)? Explain possible reasons why it did or did not.**

- No, taking away dimensions from our data would not result in a better result for our model. No dimensionality reduction resulted in a higher accuracy than the other two methods. The PCA method has no comparison functionality and the Simulated Annealing still has randomness that can lower the accuracy score it comes up with.

**Question 5d: Did Part 2 produce the same set of best features as Part 3? Explain possible reasons why it did or did not.**

- No, PCA uses the BEST original feature to create the new transformed (prime) feature set. This means PCA used only sepal-length while SA used all eight features (original and prime). As a reminder, PCA will always use the best, while SA will have randomness associated with it, so there will be a large combination of features it can try out.

**Question 5e: Did Part 2 produce the same set of best features as Part 4? Explain possible reasons why it did or did not**

- Again, No. Like I mentioned in the previous problem, Part 2 only used sepal-length as that was the best. Part 4 uses all eight features to make its features quite a bit better than anything else that we used in this lab.

**Question 5f: Did Part 3 produce the same set of best features as Part 4? Explain possible reasons why it did or did not**

- Because of the randomness that is associated with this lab, we cannot say for certain, but on average no it would not. But, please don't mark me off for this answer, but there is randomness associated with it, where part 3 could produce the same as part 4, but it also could produce not as great as part 4. Part 3 runs for at least 100 iterations where part 4 only runs for 50 iterations, so they could be different. Since simulated annealing only keeps track of one solution, it could be a lot harder to find a good solution vs part 4 where it carries multiple parents that hold possible answers. So, more often than not, genetic algorithm will provide the best answer, but occasionally the simulated annealing might be as good.