1. Based on accuracy which dimensionality reduction method, PCA, simulate annealing, or the genetic algorithm worked the best?

Genetic Algorithm worked best with accuracy 98.68.

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

PCA – Most of the variance is concentrated on one Principal component which amounts to 92% accuracy only.

Simulated annealing – Algorithm reaches 98% accuracy often but doesn’t jump to 98.67% accuracy like genetic algorithm. It still finds a sub-optimal solution.

1. 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.

1. Dimensionality reduction (PCA) has 4 components. If we use all the four, we get the same accuracy (98%) as using none.
2. Dimensionality reduction is used when we have high number of features and to eliminate by leaving components covering less variance of original set.
3. SVM do not expect features to be independent. Advantage of PCA, along with reduced dimensions, is we can get components that are independent. Since we used SVM, this aspect did not add up to extra accuracy.
4. Did Part 2 produce the same set of best features as Part 3? Explain possible reasons why it did or did not.

No.

1. Part 2, PCA, gives us only principal components.
2. Whereas for Part 3, we used both original features and Principal components. Using combination of these, we cover more domain knowledge than principal components alone. This resulted in more accuracy – 98.67%
3. Did Part 2 produce the same set of best features as Part 4? Explain possible reasons why it did or did not.

No.

1. Part 2, PCA has less accuracy than Genetic algorithm. PCA uses original features that are interdependent.
2. Genetic algorithm converged to optimal accuracy – 98.67%. The feature set for this is a combination of both original features and principal components which has more domain information.
3. Did Part 3 produce the same set of best features as Part 4? Explain possible reasons why it did or did not.

No.

1. Part 3, Simulated annealing produced the same (best) accuracy, 98.67, in earlier iterations but did not converge to it in the end after all the iterations.
2. Due to more stochastic nature in the algorithm, compared to Genetic algorithm, it did not converge in the end.
3. On the other hand, Genetic algorithm was able to converge to optimal accuracy.