Problem 2

1. **Draw scattered graph between each symptom and the label. Discuss how each symptom is dependent on the label.**

Each plot shows how a particular symptom (such as fever, cough, or fatigue) correlates with one of the disease categories (ALLERGY, COVID, COLD, FLU).

**Example:**

**Fever vs. Disease Label:** Fever is more frequently observed in cases labeled as COVID or FLU, with fewer occurrences in ALLERGY and COLD cases. This pattern suggests that the presence of fever is a strong indicator for distinguishing between COVID/FLU and the other diseases.

**Runny Nose vs. Disease Label:** Runny nose shows a stronger association with ALLERGY and COLD cases, with fewer instances in COVID. This implies that a patient presenting with a runny nose is more likely to have ALLERGY or COLD.

**Cough vs. Disease Label:** Cough is quite prevalent in COVID and FLU cases but less common in ALLERGY. This makes cough a key symptom for differentiating COVID/FLU from ALLERGY

1. **Analyze the time complexity of your code.**

time complexity of the classification algorithm is O(n \* m \* k), where:

* n is the number of test instances,
* m is the number of training instances,
* k is the number of symptoms.

1. **Give the answer in percentage. Can you increase the percentage of ratio of correctly classified instances.**

The current accuracy of the classification algorithm is calculated by comparing the predicted labels with the true labels in the test dataset.

**Can we increase the accuracy?**  
Yes, the accuracy can be improved by employing the following strategies:

* Feature Selection: Reducing irrelevant or redundant symptoms (features) can help focus on the most significant symptoms, which improves classification performance.
* Alternative Distance Metrics: Instead of using Euclidean distance, experimenting with other distance metrics such as Manhattan distance or cosine similarity may yield better results.
* Data Preprocessing: Balancing the dataset or applying feature scaling (e.g., normalizing the symptom values) can make the classification process more effective.
* Advanced Algorithms: Instead of a simple distance-based approach, using more sophisticated models like decision trees, random forests, or neural networks could increase the classification accuracy