Overview

Dataset:

The **Breast Cancer Wisconsin dataset** will be used for (link below). Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe **characteristics of the cell nuclei** present in the image. The goal is to **classify** whether the patient is **benign on malignant** based on these features.

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
This can be explored using any of the popular classification problems such as decision trees / random forest / SVM / Naïve Bayes. The features are all numerical / categorical.

I will be going with **Neural Networks** and the goal is to build the network to get a **deeper understanding** of the architecture of the network and experiment with the **number of layers / neurons at each layer / learning rates / other hyperparameters**. I will also be analyzing the results using the confusion matrix to derive deeper results.

With regards to the architecture of the network, I would be defining a multi-layer perceptron which is a class of feedforward artificial neural network, where I will be exploring the **tanh** and **relu** activation functions to see which gives the best result. Learning occurs in the perceptron by changing connection weights after each piece of data is processed based on the error and the previous layer. As I know, this too would be a supervised learning method.

This should work very well for classifying my data set, as MLPs are efficient while handling stochastic problems. I will be looking into the following papers for gaining deeper understanding of the methodology I have opted to solve the mentioned use case.

References:

1. https://www.asprs.org/wp-content/uploads/pers/1990journal/apr/1990_apr_469-473.pdf
2. https://www.sciencedirect.com/science/article/pii/S1532046403000340
3. https://ieeexplore.ieee.org/abstract/document/97934