**Programming Test: Learning Activations in Neural**

**Networks**

**Dataset for Experiment and Results: Breast Cancer Wisconsin (Diagnostic) Data Set**

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.

**ATTRIBUTE INFORMATION:**

1) ID number  
2) Diagnosis (M = malignant, B = benign)  
3-32)

Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter)  
b) texture (standard deviation of gray-scale values)  
c) perimeter  
d) area  
e) smoothness (local variation in radius lengths)  
f) compactness (perimeter^2 / area - 1.0)  
g) concavity (severity of concave portions of the contour)  
h) concave points (number of concave portions of the contour)  
i) symmetry  
j) fractal dimension ("coastline approximation" - 1)

The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field  
13 is Radius SE and field 23 is Worst Radius.

All feature values are recoded with four significant digits.

Missing attribute values: none

Class distribution: 357 benign, 212 malignant

**ALGORITHM:** The Algorithm used is Artificial Neural Network (ANN).

**BASIC EXPLORATORY DATA ANALYSIS:**

1. Imported required Libraries(Numpy, Pandas, Matplotlib.pyplot, Seaborn, SKlearn, TensorFlow)
2. Checked basic information regarding the features such as column names, datatypes of the column names, information about the column names
3. Removed unwanted features
4. Checked missing values in the data and found that there are no missing data
5. Checked for unique values of each feature in order to get much better understanding of the dataset. Confirmed that we have only 1 categorical data feature and the rest are continuous data features.

**DATA PRE-PROCESSING:**

1. **ENCODING THE CATEGORICAL VARIABLE:** To ensure that the entire dataset is of a continuous numerical form, we will be encoding the categorial variable DIAGNOSIS and converting into a numerical form, preferably into 0s and 1s.

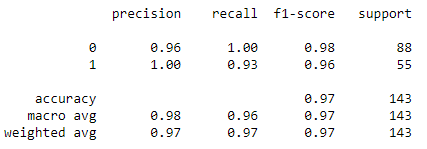
From the above table, it is clearly visible that the DIAGNOSIS feature is taking 0s and 1s as values. **[0 --> Benign; 1 --> Malignant]**

1. **SPLITTING DATASET INTO DEPENDENT AND INDEPENDENT VARIABLES:** Now finally we will be splitting the updated dataset we have into two parts. The first is a collection of the independent variables and is called the **MATRIX OF FEATURES**. The other is a collection of the dependent variables and is known as **RESPONSE FEATURE**.
2. Then used **Standard Scalar** so that the values become centered on 0 with a standard deviation 1

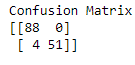
**MODEL BUILDING:**

1. Initialized the ANN(Artificial Neural Network).
2. Added the input layer and the first hidden layer using activation function “Relu”.
3. Added the output layer with activation function “Sigmoid”.
4. Compiled the ANN with optimizer = 'adam', loss = 'binary\_crossentropy' and metrics = 'accuracy'
5. Trained the ANN on the training set with validation data = (X\_test, Y\_test), batch size = 16, epochs = 100, verbose=1.
6. Predicted on the test set results

**CLASSIFICATION REPORT:**

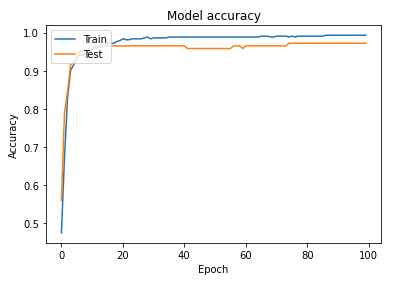


**CONFUSION MATRIX:**

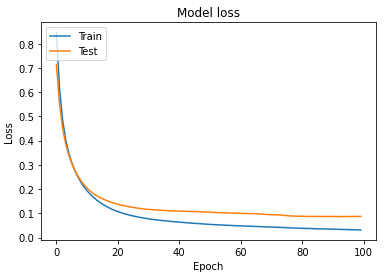


**ACCURACY SCORE:** 0.972027972027972

**Plot of Model Accuracy Vs Epoch**

****

**Plot of Mode Loss Vs Epoch**

****