Programming Test: Learning Activations in Neural Networks

Dataset Used: Breast Cancer Wisconsin (Diagnostic) Data Set

Features Present:

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)

All feature values are recoded with four significant digits.
Missing attribute values: none

ALGORITHM:

BASIC EXPLORATORY DATA ANALYSIS:

1. Imported required Libraries(Numpy, Pandas, Matplotlib.pyplot, Seaborn, SKlearn, TensorFlow)
2. Removed unwanted features
3. Searched missing values in the data.(Not Present)
4. Serched for and Confirmed that we have only 1 categorical data feature and the rest are continuous data features.

DATA PRE-PROCESSING:

1. For entire dataset to be of a continuous numerical form, we will be encoding the categorial variable DIAGNOSIS and converting into a numerical form,
   From the above datset, it is clearly visible that the DIAGNOSIS feature is taking 0s and 1s as values.
2. 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.
3. We will Standard Scalar so that the values become centered on 0 with a standard deviation 1

MODEL :

1. Initialised the ANN
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.
6. Predicted on the test set results

CLASSIFICATION REPORT:

	precision	recall	f1-score	support
0	0.96	1.00	0.98	88
1	1.00	0.93	0.96	55
accuracy			0.97	143
macro avg	0.98	0.96	0.97	143
weighted avg	0.97	0.97	0.97	143

CONFUSION MATRIX:

```
Confusion Matrix
[[88  0]
 [ 4 51]]
```

ACCURACY SCORE: 0.972