ASSIGNMENT 3

Description of the dataset:

Dataset 1(Hill-Valley): The dataset consists of 100 features (every feature is numeric) and one binary classification target(nominal). There is a total of 1212 instances. Each record shows 100 points on a 2D graph. The target is to predict if it is a Hill or Valley.

Dataset 2(steel-plates-fault): The dataset consists of 33 features (every feature is numeric) and one binary classification target(nominal). There is a total of 1941 instances. The features consist of the description of each fault, such as location, size, type of steel etc. The target is to predict if it is a common fault or other fault.

Results:

Table of Accuracy values for Hill-Valley Dataset:

Neural Network Setting	**Accuracy value**	**p-value**
No hidden layer*	0.6865126669406891	0.00294
One hidden layer with very few nodes	0.7533531963825226	0.24647
<u>One hidden layer with more nodes</u>	**0.8406833782196045**	Compared rest with this
Two hidden layers*	0.5627421796321869	0.000944

Based on the p-values, when compared to One hidden layer with many nodes, the models such as No Hidden layer and Two Hidden layers are statistically significantly different, and they are marked with *, the One hidden layer with very few nodes is not statistically significantly different.

Table of Accuracy values for steel-plates-fault Dataset:

Neural Network Setting	**Accuracy value**	**p-value**
No hidden layer	0.5522072374820709	0.07666
One hidden layer with very few nodes	0.5893391400575638	0.24358
One hidden layer with more nodes	0.6161432713270187	0.25495
<u>Two hidden layers</u>	**0.6470605343580246**	Compared rest with this

Based on the p-values above, when compared to Two hidden layers network, other networks are not statistically significantly different.

Discussion:

For the Hill-Valley dataset, adding a hidden layer to single output layer has significantly increased the accuracy. In the above table, the accuracy of the hidden layer with many nodes is more significant than any other network setting. On the contrary, the two-layer hidden network has the least accuracy. Overall, the hidden layer with many nodes is best suitable for this dataset.

For the steel-plates-fault dataset, adding a hidden layer has significantly increased the accuracy. While adding a hidden layer with more nodes has improved the accuracy too. In the above table, the accuracy of the two hidden layers is more significant than any other network setting. On the contrary to dataset 1, the two-layer hidden network has the highest accuracy. Overall, the two hidden layer network is best suitable for this dataset.