MLPR - Lab 8

Instructions:

**Duetime**: 1:30 PM

Submit Code, scatter plot for all the output images shown below as a reference.

**Step1**: Import all libraries

* Numpy
* Pandas
* Pyplot
* Self Training classifier from sklearn
* K Nearest neighbour classifier
* Confusion matrix, classification report, accuracy score, balanced accuracy score

**Step2**: Read the data given in data.csv

Step3: show a scatter plot of ‘Clump thickness vs No of week’ and plot their class labels given in column ‘Cancer stage’ (1, 2, 3, 4) for all 200 datapoints. See the reference output image below.

A graph of cancer stages

Description automatically generated

**Step4**: Now, Plot the other added 2000 datapoints given in columns ‘Clump thickness\_new’ and ‘No of week\_new’ over the previous scatter plot without using their class label. See the reference output image below.

A diagram of cancer stages

Description automatically generated

**Step5**: Now train the semi supervised model for the labeled 200 datapoints and make predictions for class label of new 2000 added datapoints and plot the scatters. Use KNN for base estimator. Use Model.fit() and model.predict() for training and prediction.

A diagram of a number of colored dots

Description automatically generated

**Step6**: Now your added 2000 new data have their predicted class and true class are given in column ‘True cancer stage’. Compute and print accuracy score, plot classification report and the confusion matrix.

A graph of confusion matrix

Description automatically generated