# AI Lab 5 : $k$-Nearest Neighbour and Perceptron

**Instruction:** In all the `.txt` files in this exercise, the following convention applies:

- $i^{th}$ row contains the data points $x_i, y_i$.
- Each $x_i \in \mathbb{R}^2$ has two coordinates $x_i(1)$ (the first entry in each row) and $x_i(2)$ (the second entry in each row).
- Each $y_i \in \{-1, 1\}$ is given in the third row.
- The *train* files contain $800$ points and the *test* files contain $200$ points.
- All the models have to be learnt using the *train* data and the performance of the models need to be evaluated on *test* data.

(Q1) [30 Marks] Build a $k$-nearest neighbour (KNN) classifier with the data given in `knn-train.txt`. For a given $k$ value, write a function `knn-acc(k)` that measures the accuracy of the KNN on the test data in `knn-test.txt` in terms of $\%$ of correct predictions. Vary $k = 1, 2, 3, \ldots$ and plot $k$ versus `knn-acc(k)`.

(Q2) [30 Marks] Learn a perceptron classifier with the data given in `perceptron-train.txt`. Write a function `prcpt-acc(k)` that measures the accuracy of the learnt perceptron on the test data in `perceptron-test.txt` in terms of $\%$ of correct predictions. For this exercise, please

1. Plot the data points, with class $+1$ in 'green', class $-1$ in 'blue' and the learnt perceptron direction $w$ in 'red' and the line $w^\top x = 0$ in 'black'.

2. Print $R$ and $\gamma$.

3. Print the number of iterations taken for perceptron to converge.

(Q3) [40 Marks] Repeat (Q2) for training data given in `perceptron-biased-train.txt` and test data in `perceptron-biased-test.txt`. Use the *padding* by 1 technique, i.e., each $x_i$ will be appended by $(x_i, 1)$, and $w = (w(1), w(2), w(3))$. Also while plotting, use the *unpadded* data for the scatter plot, and for the separating line, plot $w(1)x(1) + w(2)x(2) + w(3) = 0$ (colour coding is same as previous question).