# Homework number 1.

**Programming assignment:**
Write a program to implement the k-Nearest Neighbor algorithm. (You can use a naive search technique, you will get bonus points if you implement the more sophisticated ones.) Run the program on the data sets in `http://www.math.tau.ac.il/~mansour/ml-course`. (You should run on at least two data sets, `iris` and another one.)
*File format:* For each database you have a data file and an information file, that describing the input. In a nutshell, each line in the data file is a record, and the last value in the line is the one you would like to predict.
*Methodology:* You take a random part of the input to be a *tarining set*, this will be used in order to predict the classification. You take a different random part to be the *test set*. For each point in the test set, you compute the k-NN prediction, based on the training set, and compare it to the true value. (This is how you get an estimate to the number of errors.) The size of the taining set and $k$ should be parameters that you can adjust.

Answer the following questions:

1. What happens when you change the parameter $k$. (Increse $k$ from one till its maximum value.)
2. What happens when you increase the number of points in the training set.
3. How do the two normalization methods compare.

**Theory question:**
Suppose you have two dices. A *white dice* that has five sides color white and one black and an *equal dice* that has three sides color white and three color black. You chose a dice, the white dice has probability a third and the equal dice has probability two-thirds. You perform an experiment, you role the dice that you chose once and it comes up white.

1. What is the Maximum Likelihood dice?
2. What is the Maximum Aposteriori dice?
3. What is the probability that this is the white dice?
4. Suppose you role the *same* dice again, what is the probability it will come up white?

**The homework is due in two weeks**