kNN

THEORY

1) It is a machine learning algorithm under supervised learning which can be used for both regression and classification tasks. It is a non-parametric algorithm.
2) Minkowski distance is given as:

$$D(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \tag{1}$$

   $p = 1 \Rightarrow$ Manhattan distance $p = 2 \Rightarrow$ Euclidean distance
3) kNN normality assumption: Close data points fall into same class.
4) It finds k nearest neighbours to the data point x, the ouput label with the highest occurence among these k points is the predicted output label for x.
5) S_x is defined as:

$$S_x \subseteq D \; s.t. \; |S_x| = k \; \& \; \forall \, (x', y') \in D \;\; S_x \tag{2}$$

$$d(x, x') \geq \max_{(x'', y'') \in S_x} d(x, x'') \tag{3}$$

$$h(x) = \text{mode}(y'' : (x'', y'') \in S_x) \tag{4}$$

QUIZ

1) What is meany by chebyshev distance?
2) Overfitting occurs for:
   a) k=1
   b) k=n
3) Underfitting occurs for:
   a) k=1
   b) k=n