

Machine Learning & Machine Learning (extended)

Practice Exercise Sheet – kNN Classification

Question 1: Consider the following data set with two input attributes x and y (i.e. the coordinates of the points) and one binary output t (taking values $+$ or $-$). We want to use k -nearest neighbours (K-NN) with Euclidean distance to predict t .

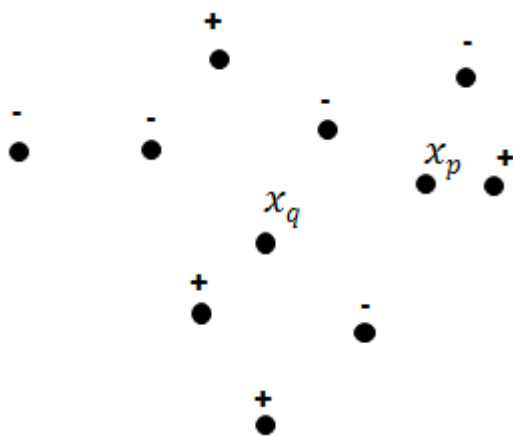
+		+		-		-
	-				-	
+		+		-		-

- Calculate the leave-one-out cross-validation error of 1-NN on this data set.
- Calculate the leave-one-out cross-validation error of 3-NN on this data set.
- Describe how you would choose the number of neighbours K in K-NN in general.

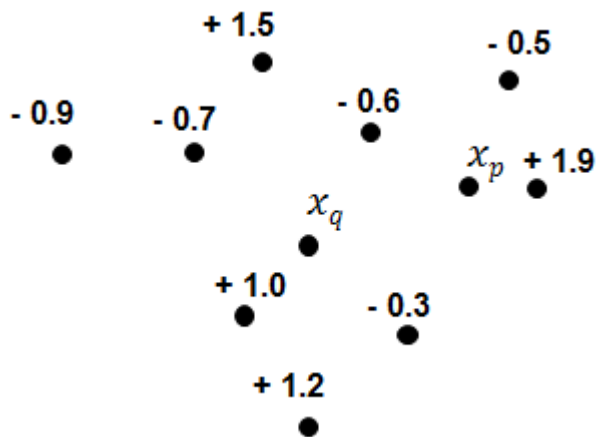
Question 2: Which of the following increases the likelihood of over-fitting and/or under-fitting? Why?

- increasing the number of neighbours k in kNN?
- decreasing the number of neighbours k in kNN?

Question 3: Assume a Boolean target function (i.e. binary classifier) and a two dimensional instance space. Determine how the kNN would classify the test instances x_p and x_q for $k = 1$, $k = 3$ and $k = 5$.



Question 4: In the diagram below, the numbers refer to the values taken by a real-valued target function. Calculate the values predicted for the target function at the test points x_p and x_q by kNN, with $k = 1$, $k = 3$, and $k = 5$.



Question 5: How do you compare weighted kNN classifier with a regular non-weighted kNN classifier? Has one of them any advantages (or disadvantages) over the other?