Ariel University

Machine Learning

Homework 4

1. Implement k-nearest neighbor on the “two_circle” data set.
   a. Sample 75 training points from the set. The remaining points are the test set.
   b. For each of k=1,3,5,7,9 and p=1,2,∞, evaluate the k-NN classifier on the test set, under the l_p distance. (The base set of the classifier is the training set.) Compute the classifier error on the test set.
   c. Repeat steps (a) and (b) 100 times, and print the average empirical and true errors for each k and p.

Which parameters of k,p are the best? How do you interpret the results? Hand in the printout and answers to these two questions in a separate file from the code.

2. Prove that the JL-transform preserves dot products up to an additive error of ±ε:

   For a set **S** of normalized vectors, we showed that the JL-transform **f** into $O(\varepsilon^{-2} \log(|S|))$ dimensions ensures that

   $(1-\varepsilon)\ \|v\| \le \|f(v)\| \le (1+\varepsilon)\ \|v\|$ for all **v** in **S**

   and

   $(1-\varepsilon)\ \|v-w\| \le \|f(v-w)\| \le (1+\varepsilon)\ \|v-w\|$ for all **v,w** in **S**

   Now prove that any embedding that satisfies the above properties also satisfies that for some constant **c**,

   $v\cdot w - c\varepsilon \le f(v)\cdot f(w) \le v\cdot w + c\varepsilon$ for all **v,w** in **S**

   What value do you get for **c**?

   Hints: Since the JL-transform is satisfied by a linear embedding, we can assume that $\|f(v-w)\| = \|f(v)-f(w)\|$.
   Also, $\|v-w\|^2 = \|v\|^2 - 2v\cdot w + \|w\|^2$.