Ariel University

Machine Learning

Homework 3

1. Implement k-nearest neighbor on the HC Temperature data set.
    a. Sample 65 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) 500 times, and print the average error for each k and p.

   Which parameters of k,p are the best? Do you see overfitting? Hand in code, printout, and answers to these two questions.

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

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

   $$(1-\varepsilon)\,\|v\| \le \|f(v)\| \le (1+\varepsilon)\,\|v\| \qquad \text{for all } \mathbf{v} \text{ in } \mathbf{S}$$

   and

   $$(1-\varepsilon)\,\|v-w\| \le \|f(v-w)\| \le (1+\varepsilon)\,\|v-w\| \qquad \text{for all } \mathbf{v,w} \text{ in } \mathbf{S}$$

   Now prove that for some constant **c**,

   $$v\cdot w - c\varepsilon \le f(v)\cdot f(w) \le v\cdot w + c\varepsilon \qquad \text{for all } \mathbf{v,w} \text{ in } \mathbf{S}$$

   Hints: Since the JL-transform is linear, $\|f(v-w)\| = \|f(v)-f(w)\|$.
   Also, $\|v-w\|^2 = \|v\|^2 - 2v\cdot w + \|w\|^2$.